Quantification and Localization of Molecular Hydrophobicity

Vom Fachbereich Chemie
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigte
Dissertation

vorgelegt von

Dipl.-Ing. Robert Jäger
aus Böblingen

Berichterstatter:	Prof. Dr. Jürgen Brickmann
Mitberichterstatter:	Prof. Dr. Hans Jörg Lindner
Tag der Einreichung:	6. Oktober 2000
Tag der mündlichen Prüfung:	20. November 2000

Darmstadt 2000

D 17

FÜR JANE, JOHANNA UND MEINE FAMILIE

Die vorliegende Arbeit wurde am Institut für Physikalische Chemie der Technischen Universität Darmstadt unter der Leitung von Prof. Dr. J. Brickmann in der Zeit von Oktober 1996 bis November 2000 durchgeführt.

Diese Dissertation wurde in englischer Sprache angefertigt. Eine Zusammenfassung in deutscher Sprache befindet sich in Kapitel 1.

Danksagung

Herrn Prof. Dr. Jürgen Brickmann möchte ich ausdrücklich danken. Zum einen für die Einführung in das Gebiet der computergestützten Chemie und der damit verbundenen Thematik dieser Dissertation, zum anderen für die vielen Freiheiten und förderlichen Maßnahmen während meiner Tätigkeit in seiner Arbeitsgruppe.

Mein besonderer Dank gilt ebenso Herrn Dr. Stefan M. Kast. Sein stetiges Engagement und Interesse an meiner Arbeit waren für mich eine große Unterstützung. Darüber hinaus war es für mich von großem Wert, von zahlreichen, bildenden Diskussionen jenseits meiner Dissertation und vielen persönlichen Gesprächen zu profitieren.

Bei Herrn Dipl.-Ing. Marco Müller möchte ich mich nicht nur für ein hervorragendes Arbeitsverhältnis bedanken, sondern vor allem für die Gelegenheit, auch die Stunden außerhalb der Universität in guter Erinnerung zu behalten.

Ebenfalls gilt mein Dank Herrn Dr. Thomas Exner und Herrn Dipl.-Ing. Matthias Keil, sowohl für Diskussionen konzeptioneller Art als auch insbesondere für die unermüdliche Hilfe zur Lösung unzähliger technischer Probleme.

Der Arbeitsgruppe von Herrn Prof. Dr. J. Brickmann, insbesondere Herrn Dr. Hans-Jürgen Bär, Herrn Dr. Dirk Zahn, Herrn Dipl.-Ing. Friedemann Schmidt und Herrn Dipl.-Ing. Bernd Schilling möchte ich für die freundschaftliche und gute Zusammenarbeit danken. Die stetige Diskussions- und Hilfsbereitschaft waren ein Garant für das Gelingen dieser Arbeit.

Inhaltsverzeichnis

1.	**ZUSAMMENFASSUNG**	**6**
1.1.	**Einleitung**	**6**
1.2.	**Das *„molecular free energy surface density“*-(MolFESD-)Konzept**	**9**
1.3.	**Parameterisierungs-Strategie**	**10**
1.3.1.	**Modell I**	**11**
1.3.2.	**Modell II**	**12**
1.3.3.	**Modell III**	**13**
1.4.	**Anwendung**	**14**
1.5.	**Literatur**	**17**
2.	**INTRODUCTION**	**19**
2.1.	**Problem**	**20**
2.2.	**Organization**	**21**
2.3.	**Literature**	**22**
3.	**LOCALIZATION AND QUANTIFICATION OF HYDROPHOBICITY: THE MOLECULAR FREE ENERGY DENSITY (MOLFESD) CONCEPT AND ITS APPLICATION TO THE SWEETNESS RECOGNITION**	**23**
3.1.	**Summary**	**23**
3.2.	**Introduction**	**24**
3.3.	**Free energy of solvation as a function of the solvent exposed surface**	**28**
3.4.	**Ionic and strong polar solutes**	**28**
3.5.	**Nonpolar solutes**	**31**
3.6.	**The empirical molecular free energy surface density (MolFESD) concept**	**34**
3.7.	**Quantification of hydrophobic contributions to the sweetness recognition**	**39**
3.8.	**Summary and Conclusions**	**48**
3.9.	**References**	**51**
4.	**FAST PREDICTION OF HYDRATION FREE ENERGIES FROM MOLECULAR INTERACTION FIELDS**	**55**
4.1.	**Abstract**	**55**

4.2.	**Introduction**	**55**
4.3.	**Statistical Basis: Three-dimensional free energy of hydration density**	**57**
4.4.	**Model I: The GRID approximation to the 3D-FED**	**58**
4.5.	**Model II: Expansion of the 3D-FED in surface-based functions**	**60**
4.5.1.	**Parameterization**	**63**
4.5.1.1.	**Three-dimensional structures**	**63**
4.5.1.2.	**Molecular surface and patch generation**	**63**
4.5.2.	**Least squares fit**	**65**
4.6.	**Final model: Simultaneous use of local and global information**	**68**
4.6.1.	**Iteration cycle**	**70**
4.6.2.	**Results**	**70**
4.7.	**Conclusions**	**74**
4.8.	**Literature**	**75**
5.	**PARAMETERIZATION STRATEGY FOR THE MOLFESD CONCEPT: QUANTITATIVE SURFACE REPRESENTATION OF LOCAL HYDROPHOBICITY**	**77**
5.1.	**Abstract**	**77**
5.2.	**Introduction**	**77**
5.3.	**The MolFESD Concept**	**79**
5.4.	**Model I: The *Grid* approximation to the 3D-FED**	**81**
5.5.	**Model II: Expansion of the 3D-FED in surface-based functions**	**84**
5.6.	**Final model: Simultaneous use of local and global information**	**87**
5.6.1.	**Iteration cycle**	**88**
5.6.2.	**Results**	**88**
5.7.	**Surface representation of the final model by means of a molecular free energy surface density**	**96**
5.7.1.	**Surface representation**	**97**
5.7.2.	**Application to the relative sweetness of a series of sucrose derivatives**	**98**
5.7.3.	**Patch-based approach**	**99**
5.7.4.	**Continuous representation**	**101**
5.8.	**Conclusion**	**104**
5.9.	**Literature**	**106**

1. Zusammenfassung

1.1 Einleitung

Die Assoziation relativ unpolarer Verbindungen in wäßriger Phase spielt für viele Gebiete der Chemie eine essentielle Rolle. Dieses Phänomen wird als *hydrophobe Wechselwirkung* bezeichnet und ist z.B. bei der Aggregation oberflächenaktiver Substanzen, der Koagulation oder für die Funktion von Detergentien von entscheidender Bedeutung. Weiterhin kann die Wirkung hydrophober Wechselwirkungen unter physiologischen Bedingungen beobachtet werden, so z.B. bei der Tertiärstruktur von Proteinen, der Stabilität biologischer Membranen oder bei molekularen Erkennungsprozessen, insbesondere für molekulare Wirt-Gast-Systeme. Eng verwandt mit der hydrophoben Wechselwirkung ist die schlechte Löslichkeit unpolarer Substanzen in Wasser, die durch den Begriff der *hydrophoben Hydratation* beschrieben wird; gemeinsam bilden sie die beiden Bestandteile, die unter der Bezeichnung *hydrophobe Effekte* zusammengefaßt werden.[1]

Bedingt durch die oben erwähnte Anwendungsbreite hydrophober Effekte ist deren thermodynamisches Verständnis auf quantitativer Ebene ein wesentlicher Fokus sowohl theoretischer als auch experimenteller Arbeiten.[1] Im Falle der hydrophoben Hydratation führte dies zu einer anhaltenden, kontroversen Diskussion,[2-8] aus der jedoch bis zum jetzigen Zeitpunkt keine einheitliche Theorie hervorgegangen ist. Zur quantitativen Beschreibung hydrophober Wechselwirkungen dient in vielen Fällen die Kenntnis experimentell bestimmter Freier Transferenthalpien. Hierbei wird der Verteilungskoeffizient P einer Substanz in einem 2-Phasen-System (meist Wasser und 1-Octanol) bestimmt, der unter Ausnutzung der folgenden Gleichgewichtsbeziehung die Freie Transferenthalpie, $\Delta G_{\text{transfer}}$, liefert

$$\Delta G_{\text{sol}}(\text{Wasser}) - \Delta G_{\text{sol}}(\text{Oktanol}) = \Delta G_{\text{transfer}} = -2.303RT \log P . \qquad \textbf{(1)}$$

Die Transferenthalpie ergibt sich aus der Differenz der individuellen Solvatationsenthalpien einer Verbindung für Wasser, ΔG_{sol}(Wasser), bzw. 1-Octanol,

ΔG_{sol}(Octanol). Gegeben durch die Gaskonstante R und die Temperatur T besteht ferner eine Proportionalität zum dekadischen Logarithmus des Verteilungskoeffizienten, log P, der experimentell vielfältig und routinemäßig zugänglich ist.[9] Neben einigen technischen Vorteilen ist es besonders die Modellfunktion des 1-Octanols für biologische Membranen (und Biophasen im allgemeinen), die es zu einem Standardlösungsmittel in Verteilungsexperimenten gemacht haben.[10-12]

Bedingt durch die beträchtliche Verfügbarkeit experimenteller log P-Daten[13] gibt es eine Vielzahl von Verfahren, meist fragment-basierter Natur, die zur Vorhersage unbekannter log P-Werte herangezogen werden.[14,15] Die schnelle und genaue Berechnung von log P-Werten ist in zweierlei Hinsicht von Bedeutung: Zum einen dient der log P-Wert als eins von insgesamt fünf Kriterien zur Abschätzung, ob ein gegebenes Molekül als potentieller Wirkstoff klassifiziert werden kann,[16] zum anderen ist er ein bedeutender Deskriptor in zahlreichen quantitativen Struktur-Wirkungsbeziehungs-(***Q****uantitative* ***S****tructure* ***A****ctivity* ***R****elationship*, QSAR-)Modellen.[17,18]

Die Kenntnis von log P-Werten, deren Berechnung auf einer zweidimensionalen (2D) Repräsentation der molekularen Struktur basiert, reicht nicht aus, um Aussagen über hydrophobe Bindungsanteile im Rahmen des gezielten Wirkstoffdesigns zu treffen. Diese 2D-Darstellung der molekularen Struktur beschränkt sich auf die An- oder Abwesenheit bestimmter Atome oder Fragmente, die anhand ihrer molekularen Nachbarschaft identifiziert werden und im Rahmen eines gegebenen Fragmentschemas einen bestimmten Typ zugewiesen bekommen. Eine Unterscheidung von Isomeren leisten nur wenige Modelle zur Vorhersage von log P.[14,15] Unter Verwendung der dreidimensionalen (3D) Molekülstruktur wurden Methoden entwickelt, die auf dem Konzept der molekularen Felder aufbauen und allgemein als 3D-QSAR-Verfahren[19,20] bezeichnet werden. Solche molekularen Felder können unter Berücksichtigung z.B. elektrostatischer, sterischer oder hydrophober Wechselwirkungen erzeugt werden, wobei hydrophobe Felder nur qualitativ zu verstehen sind und sich aus ihnen keine Vorhersage von Verteilungskoeffizienten mehr treffen läßt.[21] Die Berücksichtigung flexibler Molekülgeometrien erweitert das bisherige Verfahren und wird in der Literatur als 4D-QSAR Methodik bezeichnet.[21,22]

Mit Hilfe einer molekularen Probe können elektrostatische und sterische Felder durch Abtasten eines 3D-Gitters, das die Zielstruktur umschließt, generiert werden.

Ein diesem Verfahren zugrunde liegendes Kraftfeld wird verwendet, um für jeden Gitterpunkt eine Wechselwirkungsenergie zwischen Probe und Zielstruktur zu berechnen. Dank seiner Vielfalt an molekularen Proben und seines verläßlichen, empirischen Kraftfeldes hat sich das Programm *Grid*[23] zur Standardsoftware für die Berechnung molekularer Wechselwirkungsfelder etabliert. Breite Anwendung hat dieses Verfahren im Rahmen von CoMFA-(***co****mparative* ***m****olecular* ***f****ield* ***a****nalysis*) Studien[24,25] gefunden. Hierbei dient eine Reihe strukturell verwandter Moleküle zur Generierung molekularer Felder, deren Wechselwirkungsenergien in einer entsprechend hoch-dimensionalen Matrix abgelegt werden. Durch den Einsatz multivariater Regressionsverfahren[26] gelingt es, Unterscheidungsmerkmale der Molekülreihe in Hinblick auf das Bindungsverhalten gegenüber einem potentiellen Rezeptor abzuleiten und dadurch Bindungseigenschaften des Rezeptors zu charakterisieren. Eine Auflistung zusätzlicher Felder, die in der CoMFA-Analyse eingesetzt werden, sowie vergleichbare 3D-QSAR-Techniken und deren Limitationen finden sich in dem Übersichtsartikel von Oprea und Waller.[20]

Die Erzeugung hydrophober Felder ist eng verbunden mit dem Konzept des *molecular lipophilicity potentials* (MLP).[21] Grundlage eines solchen „Potentials“ bilden fragmentbezogene, partielle log *P*-Beiträge, die anhand experimenteller Daten durch Regressionsverfahren ermittelt wurden und in ihrer Summe den log *P*-Wert ergeben.[14] Eine räumliche Darstellung wird durch das Superponieren aller Fragmentbeiträge an jedem Punkt des Raums erreicht. Dabei wurden verschiedene Abstandsfunktionen gewählt, um Beiträge von nahen Fragmenten stärker zu gewichten.[14,27] Zwar gelingt es auf diesem Wege ein qualitatives, räumliches Bild der Lipophilie zu erzeugen, die Information über den Verteilungskoeffizienten läßt sich aus dem so generierten Feld nicht mehr ableiten. Erst eine entsprechende Darstellung in Form einer Freien Energiedichte auf der molekularen Oberfläche hat dazu geführt, daß der log *P* Wert durch Oberflächenintegration zurückerhalten werden kann.[28-30] Eine ausführliche Beschreibung des Verfahrens von Brickmann et al.[28,29] befindet sich im nächsten Kapitel.

1.2 Das „molecular free energy surface density“-(MolFESD-)Konzept

Die Berechnung der Freien Transferenthalpie gelingt bei diesem Verfahren durch Oberflächenintegration über eine kontinuierliche Funktion $\rho(\mathbf{r}_s)$ der Oberflächenkoordinaten $\mathbf{r}_s$. Diese Funktion ist benannt als *molecular free energy surface density* (MolFESD)[28-30] und repräsentiert die Freie Energiedichte eines gelösten Teilchens pro Oberflächeneinheit. Es konnte gezeigt werden, daß für die beiden Grenzfälle der Solvatation, d.h. ionischer und stark polarer bzw. unpolarer Verbindungen, der elektrostatische bzw. unpolare Anteil der Freien Solvatationsenthalpie streng als Funktion der Oberfläche einer ausgebildeten Kavität formuliert werden kann.[30] Somit kann auch für alle intermediären Fälle die Existenz einer MolFESD postuliert und zur Darstellung lokaler Hydrophobie in quantitativer Weise eingesetzt werden.

Die bisherige Modellierung der MolFESD basiert auf einem atomaren Fragmentsystem,[31] das zum einen eine Klassifizierung der molekularen Struktur erlaubt und zum anderen für jeden definierten Strukturtyp über einen partiellen log *P*-Beitrag verfügt. Diese partiellen Beiträge werden durch ein statistisches Verfahren auf eine Oberflächeneinheit bezogen.[32] Dies gelingt unter Verwendung einer normierten Fermi-Funktion[28,29] welche eine kontinuierliche Abbildung auf der molekularen Oberflächen in der Form erlaubt, daß der log *P*-Wert der betrachteten Substanz durch die Berechnung des Oberflächenintegrals ermittelt werden kann. Hieraus ergeben sich Konsequenzen, sowohl in qualitativer als auch quantitativer Hinsicht. So ermöglicht z.B. eine farb-kodierte Darstellung auf der Oberfläche die Identifikation molekularer Regionen mit hoher Affinität zu einer lipophilen Phase und erlaubt dadurch die qualitative Analyse einer gegebenen Ligand-Rezeptor Bindungswechselwirkung.[33]

Neben diesem qualitativen Aspekt ist es mit Hilfe des MolFESD-Verfahrens auch möglich, Molekülfragmente bzw. deren aufintegrierte Teiloberflächen mit physikalischen Observablen auf quantitativer Ebene zu korrelieren. Dies konnte am Beispiel einer Reihe von Saccharosederivaten und deren relativer Süße erfolgreich demonstriert werden.[30] Es wurde angenommen, daß die hydrophobe Bindungsstelle des Süßerezeptors den Fructosering der Saccharose weitgehend umschließt, so daß nur der Oberflächenanteil der Fructofuranose für die Korrelation berücksichtigt

wurde. Ferner wurde vorausgesetzt, daß die gemessene Süßkraft der halogenierten Derivate proportional zur Bindungskonstanten K ist.[34] Bedingt durch verschiedene Substitutionsmuster der Hydroxylgruppen durch Halogenatome an den Positionen 1´, 4´ und 6´ der Fructofuranose ändert sich die jeweilige Freie Energiedichte pro Oberflächeneinheit und somit das zugehörige Oberflächenintegral. Eine Korrelation dieser so bestimmten partiellen Freien Transferenthalpien mit $\ln K$ ergab einen Korrelationskoeffizienten von 0.94 und eine Fehlerquadratwurzel von 0.81.[30]

In einem neuen Ansatz zur Modellierung der MolFESD wurde in dieser Arbeit versucht, eine Parameterbasis zu schaffen, die konzeptionell auf einen geringeren Umfang experimenteller Daten angewiesen ist. Hierzu wurde neben globaler Information ($\log P$ Daten) auch lokale Information in die Parameterisierung mit einbezogen, um die Plausibilität einer lokalen Darstellung hydrophober Regionen weiter zu erhöhen.

1.3 Parameterisierungs-Strategie

Die Basis für eine Parameterisierungs-Strategie, die den oben erwähnten Anforderungen gerecht wird, ergibt sich aus der dreidimensionalen Freien-Energie-Dichte (3D-FED). Diese Größe hat ihren Ursprung in der statistisch-thermodynamischen Dichtefunktional Theorie,[35] und bildet wiederholt den Ausgangspunkt für Modellableitungen und theoretische Ansätze zum Verständnis von Solvatationseffekten.[36-38] Die 3D-FED, deren Volumenintegral über den gesamten Raum die Freie Solvatationsenthalpie wiedergibt, ist grundsätzlich zugänglich durch z.B. molekulardynamische Simulation[39] oder Integralgleichungstheorie.[40]

Dem hier beschriebenen Modell der Freien Solvatationsenthalpie liegt zunächst eine Unterteilung in wechselwirkende und auf die Ausbildung einer Kavität zurückzuführende Anteile zugrunde.[36] Die Approximation der wechselwirkenden Beiträge gelang durch Erzeugung geeigneter molekularer Wechselwirkungsfelder, während Kavitätsanteile durch eine Linearkombination oberflächen- und volumenproportionaler Terme beschrieben werden. Dieses Verfahren wurde angewandt sowohl zur raschen Vorhersage Freier Hydratationsenthalpien[41] als auch Freier Transferenthalpien.[42]

Die Approximation des wechselwirkenden Anteils der 3D-FED geschah für beide Anwendungen durch die Verwendung des Programms *Grid*[23] (siehe oben), da

hier neben rein energetischen Anteilen auch entropische Beiträge Berücksichtigung finden. Die von *Grid* zur Verfügung gestellte Wasserprobe wurde für die Modellbildung im Falle der Freien Hydratationsenthalpie, ΔG_{hyd}, benutzt. Zur Beschreibung der Freien Transferenthalpie, $\Delta G_{transfer}$, diente ein Differenzfeld, bei dem die Wechselwirkungsenergien der *hydrophoben* Probe abgezogen wurden von den Energiewerten der Wasserprobe. Unter Beibehaltung des Konzepts der molekularen Oberflächen, wurde eine Strategie entwickelt, mit der sich der wechselwirkende Anteil der 3D-FED durch eine Linearkombination oberflächenbasierter Basisfunktionen rekonstruieren läßt.

Mit Hilfe experimenteller Daten wurden für diese drei Terme, wechselwirkender Anteil der 3D-FED, Oberfläche und Volumen, Vorfaktoren bestimmt, um eine optimale Korrelation zwischen experimentellen und berechneten Freien Solvatationsenthalpien zu erreichen. Hierfür standen zwei Trainingssätze zur Verfügung: 81 Moleküle mit bekannten experimentellen Freien Hydratationsenthalpien bzw. 400 Verbindungen mit experimentell bestimmten log *P*-Werten. Beiden Applikationen wurden vorbereitende Untersuchungen vorangestellt, die der Modellvalidierung dienten.

1.3.1 Modell I

Zunächst wurde die Approximation des Wechselwirkungs-Anteils der 3D-FED durch ein geeignetes molekulares Feld überprüft. Dazu wurden die durch *Grid* erzeugten, attraktiven Wechselwirkungsenergien über das gesamte Gitter summiert und die molekularen Oberflächen[43] und Volumina der Moleküle berechnet. Die anschließende Korrelation experimenteller und berechneter Freier Solvatationsenthalpien ergab für den ΔG_{hyd}-Datensatz einen Korrelationskoeffizienten von R^2=0.92 und eine Standardabweichung von σ=1.15 kcal/mol.[41] Im Falle der Freien Transferenthalpie fielen die Ergebnisse mit R^2=0.81 und σ=1.64 kcal/mol etwas schlechter aus, was mit der Größe des Datensatzes und der damit verbunden erhöhten strukturellen Diversität zusammenhängt.[42] Dennoch validieren beide Ergebnisse die molekularen Feldapproximation. Bereits an dieser Stelle konnte festgestellt werden, daß eine Vernachlässigung des Volumenterms für den $\Delta G_{transfer}$-Datensatz das Ergebnis der Korrelation nur unwesentlich beeinflußte (R^2=0.80, σ=1.67 kcal/mol). Dieser Aspekt wird weiter unten noch einmal aufgegriffen.

1.3.2 Modell II

Für diese Modellentwicklung wurde der Versuch unternommen, die durch *Grid* generierten molekularen Wechselwirkungsfelder auf wenige, oberflächengestützte Parameter abzubilden. Hierzu wurden die erzeugten Oberflächen in diskrete Regionen aufgeteilt, die im weiteren als *Patches* bezeichnet werden. Die Aufteilung erfolgte derart, daß jeder Oberflächenpunkt einem bestimmten Atom, nämlich dem mit geringsten Abstand, zugeordnet wurde. Oberflächenpunkte, die demselben Atom zugeordnet sind, bilden einen Patch. Gemäß eines ausgewählten atomaren Fragmentsystems[31] ist jedem Atom ein entsprechender Atomtyp des Klassifikationsschemas zugeordnet. Dieser Atomtyp wird auch dem zugehörigen Patch assoziiert, so daß die Anzahl unterscheidbarer Patches durch die Anzahl verschiedener Atomtypen gegeben ist.

Desweiteren wurde für jeden Patch ein Mittelpunkt (Referenzpunkt) sowie ein dazugehöriger Normalenvektor bestimmt. Die Freie Energiedichte eines Gitterpunkts wird dargestellt als Linearkombination abstands- und winkelabhängiger Basisfunktionen, deren Aufpunkte durch die Referenzpunkte der Patches gegeben sind. Während die Abstandsfunktion durch einen Exponentialterm beschrieben wird, werden Winkelabhängigkeiten durch Entwicklung in Legendre-Polynome berücksichtigt.

Um die Anzahl der Datenpunkte für die durch lineare Regression zu bestimmenden Oberflächenparameter zu reduzieren, wurde für jeden Patch ein Würfel konstruiert, dessen Mittelpunkt die selben Koordinaten besitzt wie der Referenzpunkt des Patches. Über die Seitenlänge eines solchen Würfel kann die Anzahl der für den Fit relevanten Gitterpunkte gesteuert werden, die offensichtlich innerhalb des Würfels liegen müssen. Optimale Kantenlängen wurden zu 2.5Å (Transfer) bzw. 3.5Å (Hydratation) bestimmt. Zusätzlich wurden nur Gitterpunkte berücksichtigt, deren Freie Energiedichten auf attraktive (statt repulsive) Wechselwirkung zurückzuführen sind und deren Winkel zum Normalenvektor des Patches weniger als 110° beträgt.

Die Anpassung an experimentelle Freie Solvatationsenthalpien erfolgte in zwei Schritten. Durch lineare Regression wurden zunächst die Koeffizienten der oben beschriebenen Linearkombination oberflächenbasierter Basisfunktionen bestimmt, um eine optimale Abbildung der räumlichen Energiedichten zu erhalten. An dieser Stelle tragen ausschließlich *lokale* Informationen, gegeben durch die approximierte

3D-FED, zur Modellentwicklung bei. Das Ergebnis dieser Regression ist exzellent, sowohl für die Freien Hydratationsenthalpien[41] (R^2=0.92, σ=0.27 kcal/mol) als auch für die Freien Transferenthalpien (R^2=0.89, σ=0.60 kcal/mol).[42] Diese Ergebnisse beziehen sich auf die Entwicklung der Legendre Polynome bis zur 2. Ordnung, d.h. die Basisfunktionen beinhalten einen Kosinusterm. Höhere Ordnungen verbessern die Fit-Statistik nur geringfügig, so daß im folgenden nur die Resultate der 2. Ordnung verwendet werden.

Der zweite Schritt verläuft gemäß Modell I unter Verwendung *globaler* Information, diesmal mit der eben beschriebenen Approximation des Wechselwirkungsanteils der Freien Solvatationsenthalpie (siehe oben). Das Ergebnis dieser Korrelation ist von derselben Güte, die bereits in Modell I beobachtet werden konnte. Für den ΔG_{hyd}-Datensatz ergibt sich R^2=0.88 und σ=1.38 kcal/mol, während für die 400 Moleküle des $\Delta G_{transfer}$-Datensatzes R^2=0.83 und σ=1.54 kcal/mol erhalten wurden. Die Darstellung des Freien Energiedichtefeldes anhand weniger oberflächengestützter, strukturbezogener Parameter ist adäquat (vgl. Modell I) in der Lage, Freie Solvatationsenthalpien zu reproduzieren und legitimiert somit Modell II. Ebenso konnte bestätigt werden, daß eine Vernachlässigung des Volumenterms im Falle der Freien Transferenthalpien zu keiner signifikanten Änderung der Korrelationsergebnisse führte. Aus diesem Grund wurde in der Entwicklung des endgültigen Modells (Modell III) auf einen dem Volumen proportionalen Beitrag zur Beschreibung der Kavitätsanteile verzichtet (nur für Transferenthalpien).

1.3.3 Modell III

Im vorherigen Modell erfolgte die Modellentwicklung in zwei Stufen, in denen sukzessive lokale und globale Information zur Modellanpassung herangezogen wurde. Durch Angleichen der Parameterbasis leistet Modell III eine Zusammenführung beider Schritte, so daß lokale und globale Information *simultan* regrediert werden können. Bedingt durch nicht-lineare Abhängigkeiten der Modellparameter erfolgt diese Optimierung im Rahmen eines iterativen Verfahrens. Als Stopkriterium wird die Konvergenz der Standardabweichung des verwendeten Trainingsdatensatzes herangezogen. Eine optimale Gewichtung lokaler und globaler Daten wird anhand der minimalen Standardabweichung für die Vorhersage Freier Solvatationsenthalpien eines Testsets evaluiert.

Die simultane Berücksichtigung lokaler und globaler Informationen im iterativen Optimierungszyklus konnte die Güte der Modellfunktionen entscheidend verbessern. So erreichte der Korrelationskoeffizient für die Freien Hydratationsenthalpien einen Wert von R^2=0.99 bei einer Standardabweichung von σ=0.27 kcal/mol. Die Vorhersage von ΔG_{hyd} für ein aus 10 Molekülen bestehendes Testset gelang innerhalb einer Fehlergrenze von σ=0.63 kcal/mol und lag damit deutlich unter der experimentellen Genauigkeit.

Auch im Falle der Freien Transferenthalpien konnte eine ähnliche Verbesserung der Korrelation gefunden werden (R^2=0.95, σ=0.60). Hier ergaben sich weitere Konsequenzen im Sinne einer kontinuierlichen Oberflächenfunktion (MolFESD), die im nächsten Abschnitt beschrieben werden.

1.4 Anwendung

Eine besondere Charakteristik des entwickelten Modells III besteht in der Konservierung der durch die 3D-FED gegebenen, lokalen Information. Dies gelingt durch die entsprechend angepaßten, patch-basierten Koeffizienten, die in ihrer Linearkombination den wechselwirkenden Anteil der Freien Energiedichte an jedem Raumpunkt abbilden. Diese Abbildung der 3D-FED auf wenige, oberflächengestützte Parameter rechtfertigt eine plausible Interpretation im Sinne eines lokalen Beitrags zu einer gegeben globalen Observablen. Somit ist allein unter Kenntnis dieser diskreten Koeffizienten die Behandlung von Teiloberflächen, und die damit verbundenen Berechnung partieller thermodynamischer Größen möglich.

Der Übergang zu einer kontinuierlichen Darstellung wird anhand dieser auf den Referenzpunkten der Patches lokalisierten, partiellen Parameter erreicht. Sie dienten als Stützstellen für eine kontinuierliche Verteilung auf der Molekülobefläche mittels Gauss-Funktionen. Die Division durch das einem Oberflächenpunkt zugehörige Oberflächenelement liefert eine oberflächenbezogene Darstellung des Wechselwirkungs-Anteils der Freien Transferenthalpie. Da in diesem Fall auch der Kavitätsanteil nur durch einen oberflächenproportionalen Term beschrieben wird, war es möglich, die Modellparameter zu vereinen, um zu einer einheitlichen Beschreibung analog der MolFESD (siehe oben) zu gelangen. Die Oberflächenintegration dieser so gebildeten Freien Transferenthalpiedichte liefert identische

log *P*-Werte wie Modell III; der quantitative Charakter dieser kontinuierlichen Darstellung bleibt erhalten. Zur Validierung dieser diskreten bzw. kontinuierlichen Lokalisation wurde die Applikation bzgl. der Saccharosederivate[30] erneut aufgegriffen.

Für den Oberflächenanteil der Fructofuranose wurde die Freie (partielle) Transferenthalpie gemäß der etablierten, diskreten Parameter aus Modell III berechnet und mit der relativen Süßkraft der Zuckerderivate korreliert.[42] Das exzellente Ergebnis (R^2=0.96, σ=0.64) ist einerseits Indiz für die hohe Plausibilität der lokalen, patch-basierten Parameter, anderseits belegt es die physikalische Relevanz des berechneten Teilintegrals.

Für die kontinuierliche Freie Transferenthalpie-Oberflächendichte (siehe oben) ergibt sich eine Abhängigkeit in Bezug auf die verwendete Gauss-Funktion. Je nach Verlauf der Gauss-Funktion (gegeben durch z.B. die Position des Wendepunktes) ergibt sich eine unterschiedliche, lokale Darstellung auf der Molekülobefläche. Dies führte zu leicht verschiedenen Ergebnissen für die partielle Freie Transferenthalpie, die durch Integration über den Oberflächenanteil der Fructofuranose erhalten wurde. Die beste Korrelation ergab sich für die Gauss-Funktion, deren Wendepunkt bei 2Å lag. Anhand dieses Freiheitsgrades konnte eine optimierte Darstellung bzw. farbkodierte Visualisierung der Freien Transferenthalpie im Sinne eine Oberflächendichte (MolFESD) erzielt werden. Neben der damit verbunden qualitativen Erkenntnis über hydrophobe bzw. hydrophile Molekülregionen und deren gegenseitiger Beeinflussung, wird diese Darstellung auch ihrem quantitativen Anspruch gerecht: Die Korrelation dieser durch Oberflächenintegration berechneten partiellen Freien Transferenthalpien ergab einen Korrelationskoeffizienten von R^2=0.96 bei einer Standardabweichung von σ=0.65. Das Ergebnis ist in gutem Einklang mit der statistischen Güte, die durch Benutzung der diskreten, patch-basierten Parameter erzielt wurde (siehe oben) und rechtfertigt somit das Konzept einer MolFESD.

Durch die simultane Verwendung lokaler und globaler Informationen konnte ein quantitatives Modell zur Beschreibung molekularer Hydrophobie etabliert werden, das zugleich in seiner lokalen Betrachtung höchst plausibel ist und den quantitativen Ansprüchen gerecht wird. Es ist zu erwarten, daß eine automatisierte Version dieses Modells als sinnvolle Ergänzung in QSAR-Studien einfließt, insbesondere zur Beschreibung lokaler Beiträge für die Ligand-Rezeptor Wechsel-

wirkung. In einem weiterführenden Projekt soll die Fähigkeit dieses Verfahrens bzgl. der Vorhersage Freier Hydratationsenthalpien für eine Modellentwicklung zur Bestimmung Freier Bindungsenthalpien ausgenutzt werden.

1.5 Literatur

1) Blokzijl, W.; Engberts, J.B.F.N. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1545.
2) Scheraga, H. A. Journal of Biomolecular Structure & Dynamics **1998**, 16, 447.
3) Lazaridis, T. *J. Phys. Chem. B* **2000**, *104*, 4964.
4) Ruelle, P.; Kesselring, U. *J. Pharm. Sci.* **1998**, *87*, 987.
5) Besseling, N. A. M.; Lyklema, J. *J. Phys. Chem. B* **1997**, *101*, 7604.
6) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.
7) Hummer, G. *J. Am. Chem. Soc.* **1999**, *121*, 6299.
8) Filipponi, A; Bowron, D. T.; Lobban, C.; Finney, J. L. *Phys. Rev. Lett.* **1997**, *79*, 1293.
9) Sangster, J. Octanol-Water Partition Coeffients: Fundamentals and Physical Chemistry, Vol. 2; John Wiley & Sons: Chichester 1997.
10) Tute, M. S. *Methods and Principles in Medicinal Chemistry*, Vol. 4; Wiley-VCH: Weinheim, 1996; Chapter 2.
11) DeBolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5316.
12) Hansch, C.; Dunn III, J.; *J. Pharm. Sci.* **1972**, *61*, 1.
13) *Star list Database*; Biobyte Corporation; Pomona, CA, 1997.
14) Buchwald, P.; Bodor, N. *Curr. Med. Chem.* **1998**, *5*, 353.
15) Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868.
16) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Freeny, P. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
17) Leo, A. J.; Hansch, C. Perspective in Drug Discovery and Design **1999**, 17, 1.
18) Hansch, C. *Acc. Chem. Res.* **1993**, *26*, 147.
19) Kubinyi, H. 3D QSAR in Drug Design. Theory, Methods and Application, Vol 1; Kluwer/Escom: Dordrecht, 1993.
20) Oprea, T. I.; Waller, C. L. *Reviews in Computational Chemistry*, Vol. 11; Wiley-VCH: New York, 1997; Chapter 3.
21) Carrupt, P.-A.; Testa, B.; Gaillard, P. *Reviews in Computational Chemistry*, Vol. 11; Wiley-VCH: New York, 1997; Chapter 5.
22) Vedani, A.; McMasters, D. R.; Dobler, M. *Quant. Struct.-Act. Relat.* **2000**, *19*, 149.
23) Goodford, P. *J. Chemometrics* **1996**, *10*, 107.
24) Cramer, D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
25) Kubinyi, H. Encyclopedia of Computational Chemistry;Wiley: New York, 1998.
26) Greco, G.; Novellino, E.; Martin, Y. C. *Reviews in Computational Chemistry*, Vol. 11; Wiley-VCH: New York, 1997; Chapter 4
27) Carrupt, P.-A.; Gaillard, P.; Billois, F.; Weber, P.; Testa, B.; Meyer, C.; Pérez, S. *Methods and Principles in Medicinal Chemistry*, Vol. 4; Wiley-VCH: Weinheim, 1996; Chapter 12.
28) Pixner, P.; Heiden, W.; Merx, H.; Möller, A.; Moeckel, G.; Brickmann, J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309.
29) Brickmann, J.; Jäger, R. *Analusis* **1999**, *27*, 15.

30) Jäger, R.; Schmidt, F.; Schilling, B.; Brickmann, J. *J. Comput.-Aided Mol. Design* **2000**, *14*, 631.
31) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762.
32) Jäger, R. *Diplomarbeit*, TU Darmstadt, 1996.
33) Hollstein, M; Moeckel, G.; Hergenhahn, M.; Spiegelhalder, B.; Keil, M.; Werle-Schneider, G.; Bartsch, H.; Brickmann, J. *Mutat. Res.-Fund. Mol. M.* **1998**, *405*, 145.
34) Lee, C. K. Adv. Carboh. Chem. Biochem. **1987**, 45, 199.
35) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, 1990; Chapter 6.7.
36) No, K.T.; Kim, S.G.; Cho, K.-H.; Scheraga, H.A. *Biophysical Chemistry* **1999**, *78*, 127.
37) Son, S. H.; Han, C. K.; Ahn, S. K.; Yoon, J. H.; No, K. T. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 601.
38) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.
39) Allen, M. P.; Tildesley, D. P. *Computer Simulations of Liquids*; Oxford University Press: Oxford, 1989.
40) Kollmann, P. A. *Acc. Chem. Res.* **1996**, *29*, 461.
41) Jäger, R.; Kast, S. M. *J. Phys. Chem. A*, submitted.
42) Jäger, R.; Kast, S. M.; Brickmann, J. *J. Phys. Chem. A*, in preparation.
43) Connolly, *Science* **1983**, *221*, 709.

2 Introduction

Quantitative information on solvation and transfer free energies is indispensable to the understanding and modeling of biomolecular structure and binding. The presence of nonpolar compounds or molecular fragments in an aqueous environment renders entropic contributions dominant for the assessment of the Gibbs free energy with respect to solvation or binding. Such entropically unfavorable situations are generally termed *hydrophobic effects*.[1] Quantitative treatment is desirable due to the relevance of hydrophobic effects for many processes in aqueous solution, e.g. the stability of biological membranes, the tertiary structure of proteins, molecular recognition phenomena, or transport and diffusion processes in biological systems. Besides such biological aspects, hydrophobic effects are also important for the function of detergents, coagulation, and many surface-related phenomena. Hydrophobic effects comprise two distinguishable aspects: the transfer of a nonpolar substance from the gas phase into water, or *hydrophobic hydration*, and the association of nonpolar moieties in aqueous solution, called *hydrophobic interaction*.

The hydration of nonpolar compounds is microscopically a complex process because of the unique characteristics of the water phase. Experimentally one finds a large positive free energy of hydration, a negative enthalpy and entropy contributions at room temperature, and a large heat capacity. A general, quantitative microscopic theory to explain this behavior is yet to be found and the subject of hydrophobic hydration is discussed controversially in the literature.[1,2]

The association of nonpolar moieties in aqueous solution (e.g. formation of micelles) or the transfer of a nonploar ligand from the water phase into a hydrophobic receptor pocket are examples of hydrophobic interactions. Such behavior is modeled by transfer experiments where a nonpolar substance is allowed to distribute between water and an organic, immiscible phase.[3] 1-Octanol has become the standard organic solvent to determine the partition coefficient P (usually reported as $\log P$) of a given compound. Besides certain practical benefits of 1-octanol, it resembles the structure of membrane lipids and can therefore well be used as a biophase analogue.[4,5] The transfer free energy of a substance for the 1-octanol/water system is taken as a quantitative measure for the affinity of a compound or molecular fragment to a non-aqueous solvent. Due to the extensive availability of experimental $\log P$ values, many empirical, often fragmental, approaches were developed to predict this partition coefficient.[6] As an extension, three-dimensional properties were

derived (e.g. hydrophobic fields) and efficiently used for the modeling of ligand-receptor interactions.[7]

Following this development, it is the main intention of this thesis to establish a model that allows (a) to predict the log *P* value or transfer free energy for a given compound and (b) to yield a hydrophobic profile on the molecular surface to identify regions with strong affinity to a non-aqueous environment. A parameterization strategy is presented that accounts for both global and local information simultaneously thereby yielding a plausible representation of local hydrophobicity.

2.1 Problem

For a quantitative surface-based treatment of the transfer free energy an empirical model was developed earlier.[8] This model is based on a scalar quantity, the *molecular free energy surface density* (MolFESD), defined on the molecular surface in such a way that the surface integral recovers the transfer free energy. The parameterization of this model is based on an established set of atomic incremental log *P* values[9] combined with a membership function governing the contribution of an atom to a surface region. An important aspect of this surface-based representation of the free energy of transfer is the use of partial thermodynamic properties which are derived from integration over incremental surface areas. A detailed description of the physical basis and model derivation is given in chapter 3 of this thesis. Moreover, implications and consequences are outlined when the incremental treatment of the free energy surface density is applied to the immersion depth of a compound at a given phase boundary. It is also used for the problem of finding a correlation between hydrophobic surface areas of a series of sucrose derivatives and their relative sweetness.

Clearly, the MolFESD concept is dependent on the availability of experimental data and the quality of their measurement. Despite the extensive amount of experimental log *P* data, one would desire a parameterization basis in the absence of experimental flaw or synthesis difficulties. The *three-dimensional free energy density* (3D-FED) offers a physically sound theoretical basis[10,11] towards the establishment of predictive models with limited empirical character. While this density is in principle accessible through computational approaches such as molecular dynamics or Monte-Carlo methods, in this work the program *Grid*[12] was

employed for a rapid approximation of the interaction part of the 3D-FED. Based on an empirical force field, the program *Grid*[12] determines the interaction energy of a particular probe molecule (e.g. water) for each intersection of a chosen grid which encloses the target structure.

Due to the 3D-nature of the FED, conformational characteristics of a molecule are considered in the parameterization process. Most fragmental models for the prediction of log *P* do not take into account conformational aspects.[6] Such models encode the molecular structure in a two-dimensional string and the mere presence or absence of a particular atom type determines the estimate for the log *P* value.

To complete the description of the solvation free energy cavity contributions are modeled as a linear combination of surface and volume proportional terms.[13] Experimental data is employed for an appropriate scaling of model parameters. A detailed model derivation and the successful prediction of hydration free energies are presented in chapter 4, while chapter 5 focuses on the transfer free energy. Besides the predictive capability with respect to the hydration or transfer free energy, in chapter 5 a surface-based representation is developed to derive the free energy surface density. This chapter concludes with a comparison of application results previously determined in chapter 3.

2.2 Organization

This thesis has the following structure: chapters 3, 4, and 5 are independent sections each one providing an introduction, conclusion, and literature part. This is due to the fact that these three sections have been submitted or prepared for publication. While section 3 has already been published,[14] section 4 has been submitted to the Journal of Molecular Graphics and Modelling. Section 5 presents a manuscript intended for submission.

At this point it is important to stress that chapters 3.4 and 3.5 are external contributions and do *not* represent investigation done by the author of this thesis. For these two chapters credit must be given to the co-authors of this publication, Friedemann Schmidt and Bernd Schilling. These two external parts were included in this thesis to represent the publication[14] in its full, original extent in order to avoid incompleteness thus ensuring full comprehension of the presented work.

2.3 Literature

1) Blokzijl, W.; Engberts, J. B. F. N. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1545.
2) Southhall. N. T.; Dill, K. A. *J. Phys. Chem. B* **2000**, *104*, 1326.
3) Sangster, J. Octanol-Water Partition Coeffients: Fundamentals and Physical Chemistry; John Wiley & Sons: Chichester 1997; Vol. 2.
4) DeBolt, S. E.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5316.
5) Hansch, C.; Dunn III, J.; *J. Pharm. Sci.* **1972**, *61*, 1.
6) Buchwald, P.; Bodor, N. *Curr. Med. Chem.* **1998**, *5*, 353.
7) Böhm, H.-J.; Klebe, G.; Kubinyi, H. *Wirkstoffdesign*; Spektrum, Akad. Verl.: Heidelberg, 1996; Chapter 21.
8) Pixner, P.; Heiden, W.; Merx, H.; Möller, A.; Moeckel, G.; Brickmann, J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309.
9) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762
10) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, 1990; Chapter 6.7.
11) No, K.T.; Kim, S.G.; Cho, K.-H.; Scheraga, H.A. *Biophysical Chemistry* **1999**, *78*, 127.
12) Goodford, P. *J. Chemometrics* **1996**, *10*, 107.
13) Ben-Naim, A.; Lovett, R. *J. Phys. Chem. B* **1997**, *101*, 10535.
14) Jäger, R., Schmidt, F.; Schilling, B.; Brickmann, J. *J. Comput.-Aided Mol. Design* **2000**, *14*, 631.

3 Localization and Quantification of Hydrophobicity: The Molecular Free Energy Density (MolFESD) Concept and its Application to the Sweetness Recognition

3.1 Summary

A method for the localization, the quantification, and the analysis of hydrophobicity of a molecule or a molecular fragment is presented. It is shown that the free energy of solvation for a molecule or the transfer free energy from one solvent to another can be represented by a surface integral of a scalar quantity, the molecular free energy surface density (MolFESD), over the solvent accessible surface of that molecule. This MolFESD concept is based on a model approach where the solvent molecules are considered to be small in comparison to the solute molecule, and the solvent can be represented by a continuous medium with a given dielectric constant. The transfer energy surface density for an 1-octanol/water system is empirically determined employing a set of atomic increment contributions and distance dependent membership functions measuring the contribution of the increments to the surface value of the MolFESD. The MolFESD concept can be well used for the quantification of the purely hydrophobic contribution to the binding constants of molecule-receptor complexes. This is demonstrated with the sweeteners sucrose and sucralose and various halogen derivatives. Therein the relative sweetness, which is assumed to be proportional to the binding constant, nicely correlates to the surface integral over the positive, hydrophobic part of the MolFESD, indicating that the sweetness receptor can be characterized by a highly flexible hydrophobic pocket instead of a localized binding site.

3.2 Introduction

Many concepts in chemistry are based on vaguely defined quantities. Nevertheless, they often work surprisingly well for the interpretation of experimental result and the systematic planning of new experiments. A prominent example for this statement is the concept of hydrophobicity. The interaction of a nonpolar molecule with an aqueous solvent or the quasi-attractive interaction of organic compounds or nonpolar groups in the water phase is known as the *hydrophobic hydration* or *hydrophobic interaction*, respectively.[1,2] Hydrophobic effects play a key role in many chemical phenomena in aqueous solution. The stability of biological membranes as well as the tertiary structure of proteins in aqueous solution depend critically on interactions between nonpolar moieties. In molecular recognition hydrophobic bonding is one of the most important forces that arise from the interaction of nonpolar regions of a guest molecule (a pharmaceutical drug for example) with its host molecule (the receptor). Hydrophobic effects are also important for various solution phenomena such as surfactant aggregation, mineral flotation, coagulation, complexation, and detergency.

From the point of view of phenomenological thermodynamics, the hydrophobic effect can be quantified using the solvation free energy of a certain molecular compound ΔG or the Gibbs transfer free energy change $\Delta\Delta G = \Delta G_{transfer}$ of this compound between an aqueous and an apolar liquid phase. A similar approach can be used to study the hydrophobic interaction of two molecules in aqueous solution. There are a lot of thermodynamic data available now for a rich variety of molecules including values for the heat of solvation, dissociation and association constants, partition coefficients of molecules in liquid-liquid systems, specific heat capacities etc. In many cases these quantities have been measured as functions of temperature and other thermodynamic variables over a wide range. However, up to now, there is still no physical model available which allows a clear correlation between microscopic molecular properties (like those which can be can be measured from isolated molecule experiments or quantum chemical calculations) and macroscopic (thermodynamic) data which are connected to the term hydrophobicity or hydrophobic interaction. For an extensive discussion of this topic see the excellent review of Blokzijl and Engberts.[1]

The understanding of the molecular nature behind the hydrophobic effect or at least a uniquely defined molecular scenario which allows correlation of the kind mentioned above is obviously essential for a quantitative treatment of this phenomenon. The aim of the present work is directed towards the establishment of a semi-empirical scenario in which molecular properties or properties of molecular fragments can be directly correlated to partial ΔG values, i.e. correlated to the contribution of hydrophobic interaction to the overall binding constant of a molecular complex (a substrate receptor complex for example). The strategy is to develop a concept which has a clear physical basis on one hand but wherein certain open parameters are fitted to experimental (thermodynamic) data on the other hand. The procedure should be fast enough to be used within QSAR studies or even database screening.

A standard thermodynamic measure for a quantitative comparison of the overall hydrophobicity or lipophilicity of different molecules can be obtained from the partition coefficients P[3-9] or the hydrophobic index $\log P$[10,11] of the sample in polar-apolar heterogeneous reference systems. The two phases are in equilibrium if the relation

$$\log P = -\frac{\Delta\overline{G}_{\mathrm{I,II}}}{2.303\,RT} = -\frac{\Delta\overline{G}_{\mathrm{transfer}}}{2.303\,RT} \qquad \textbf{(1)}$$

holds. The employment of the partition coefficient in QSAR studies was recently reviewed by Buchwald and Bodor.[12]

Fujita et al.[3] could prove within a comprehensive study the additive-constitutive nature of the partition coefficient. Based on their results the hydrophobicity of a molecule – measured, for example, by its hydrophobicity index ($\log P$-value) – can be regarded as the sum of increments related to the hydrophobicities of its fragments, f_i:

$$\log P = \sum_i f_i \qquad \textbf{(2)}$$

Many authors[4-11,13-20] published sets of fragmental values for partition coefficients (dominantly for the 1-octanol/water system). In place of these we refer to an empirical quantification of lipophilicity contributions of individual atoms which has

been accomplished by Ghose and Crippen.[8,9,20] In their treatment the atomic contributions were defined by using a classification of each atomic fragment according to the number and nature of their next (and second next) connected neighbors. Up to now roughly 120 constitutive atom types are listed, describing the atoms in their individual structural environment.[8-11,20] A modification and extension of this set has been elaborated in the group of the present authors.[21] The subdivision of molecular lipophilicity into fragmental contributions seems to be a prerequisite for the manifestation of local hydrophobicity, however, this scheme is not sufficient in any case. So the volume of the solute molecule, its surface area or its 3D-structure or shape can only be rudimentarily considered in such a scenario.

In a recent paper from the group of the author[22] a molecular hydrophobicity mapping (MHM) approach was suggested which was particularly designed for projecting atomic increment values on the molecular surface (a solvent accessible surface (SAS) or contact surface[23] of a molecule)[*]. The MHM approach is well suited for a qualitative discussion of local hydrophobicity on the basis of incremental contributions because the formalism induces a weighted projection of the incremental f_i-values of Crippen and coworkers[8,9] onto the molecular surface and so allows an easy visualization of this quantity. One cannot compute quantitative data like $\Delta G_{transfer}$ or the partition coefficient P from molecular hydrophobicity maps regardless of the qualitative value of the MHM. Such a quantification can also not be performed on the basis of so called *molecular lipophilicity potentials* (MLP) which have been first introduced by Audry et al.[24] and modified by a variety of authors.[25-28] Again, despite of the success of the MLP approach in CoMFA[29] and QSAR[30] there is no physical justification for the use of one or the other distance function in this “potential” while using fragmental parameters as “partial charges” located at the positions of the atoms in a molecule.[22]

The quantitative treatments in this paper base on the molecular surface concept. It has been shown[31] that there is a correlation between ΔG, ΔH and $T\Delta S$ (Δ stands for the thermodynamic process of solvation) for many apolar gases solved in water and in n-hexane and the molecular surface area. The Gibbs energies for the transfer of apolar

[*] Here and in the following text we will use the term "molecular surface" in a sense of Connolly,[23] i.e. a contact surface of "spherical" water probe (radius r=1.4 Å) with a CPK-model of a molecule of given conformation is taken as a reference

solutes from nonpolar solvents to water have also been correlated with accessible molecular surface areas and it was shown[32-44] that the Gibbs energy per surface area involved in the transfer of apolar compounds from a hydrocarbon-like solvent to water amounts to 70 to 130 $Jmol^{-1}A^{-2}$. The exact magnitude strongly depends on the reference solvent, the computational procedure used to determine the surface area, and the fitting strategy. Abraham[36,37] as well as Richards and coworkers[42,43] assumed that the transfer free energy $\Delta G_{transfer}$ for one mole substance from one solvent to the other can be composed of additive contributions which are related to the solvent accessible surface of fragments of the molecule. A similar scenario has been already suggested by Tanford.[35] The surface concept was also used in theoretical works like that of Lee[45] whose considerations were based on the assumption that the change of enthalpy and entropy are linear functions of the surface area of the solute. A promising *ab initio* based approach to localize the electrostatic part of solvation free energy on molecular surface area has recently been presented by Luque and coworkers.[48] Kellogg and Abraham[46,47] used empirical atom-based hydrophobic parameters to estimate protein-ligand binding constants. The HINT program has been proved to be useful in QSAR and CoMFA studies.

Independently from the fact that the molecular surface concept is very useful for the quantitative treatment of thermodynamic data related to the solvation process or the transfer of a solute molecule from one solvent to another, it should be noticed, that these data are in general not in quantitative agreement with the macroscopic surface tension or other surface energies.[49]

In this work we present a concept wherein the thermodynamic data ΔG, ΔH and $T\Delta S$ of solvation or transfer can be calculated as a surface integral over a quantity which has been termed *molecular free energy surface density* (MolFESD) $\rho(\mathbf{r})$, a continuous function of the surface coordinates $\mathbf{r}$ which can be considered as the free energy per unit surface of the solute. The paper is organized as follows: In the next section it is shown analytically that the enthalpic part of $\rho(\mathbf{r})$ can be obtained from electrostatic argument for those cases where the solute molecule is modeled as a charge distribution separated from a continuous dielectric solvent. Some arguments are reviewed for the assumption that $T\Delta S$ can be modeled as a surface extensive property. This is followed by a section that deals with the explicit construction of the model. It is outlined how to compose the MolFESD from atomic fragments of the solute molecule. The parameterization is based

on the partition of solute molecules in an 1-octanol/water system. We have chosen this system because there is a broad spectrum of experimental data available. It is most relevant for the hydrophobic interaction in biological systems since it has been demonstrated that hydrated 1-octanol can well serve as a hydrophobic biophase analogue.[6,50,51] In the application section we demonstrate the quantification of hydrophobic interactions as a measure for the sweet taste sensation of various compounds using the MolFESD approach. The final section contains a summary and an outlook to further applications.

3.3 Free energy of solvation as a function of the solvent exposed surface

In this section the physical basis for the introduction of the MolFESD concept is given and some limitations are outlined. We discuss the evaluation of the electrostatic interaction energy between solute and the surrounding solvent as well as the entropic contributions in terms of a surface energy density functional.

Considering solution thermodynamics of different classes of molecules, there are two limiting cases of (i.) ionic/strongly polar solutes and (ii.) nonpolar organic compounds. While the free energy of solvation, ΔG_{solv}, of the former tends to be dominated by electrostatic energy, it is the work of cavity formation that makes the dominant contribution to ΔG_{solv} of the latter. In the following we discuss how the molecular solvation energy in whatever case can be calculated as a function of the solvent-exposed molecular surface.

3.4 Ionic and strong polar solutes

Classical electrostatics provide an important tool for the calculation of solvation energies of polar and ionic solutes in aqueous solution. In the widely accepted macroscopic continuum model[52-64] the structure of solvent media is neglected. Continuum solvent models have been developed for use with both quantum mechanical and empirical force fields.[65-76] While taking an explicit

representation of solute atoms – e.g. an arbitrary charge distribution in a cavity – the average electrostatic properties of a solvent medium are represented by an infinite polarizable dielectric medium with constant permittivity ε_v. Accounting for atomic polarizability, the molecular cavity is defined as a low dielectric medium $\varepsilon_\mu << \varepsilon_v$.

By extrapolating thermodynamic properties of macroscopic systems to a microscopic level, electrostatic contributions to free energies of solvation (ΔG_e) can be calculated in terms of the difference in the total reaction field energy of solvent and vacuum upon the cavity-bound charge distribution. Electrostatic potential maps are obtained by solving the Poisson equation, which can either be done numerically[53-56,59,61-65] or analytically[77-81] for simple geometry. ΔG_e is derived by an integration or summation procedure of the electrostatic energy densities of a hypothetical solute in vacuum and in solution over all space.

An alternative two-step approach for ΔG_e is offered by calculating the transfer energy ΔG_τ[82,83] for the substitution of a homogenous and infinite dielectric medium with permittivity ε_μ with a dielectric medium ε outside the molecular cavity. ε stands either for vacuum or solvent medium. ΔG_τ can be calculated from the gradients of the electrostatic potentials $\nabla\varphi_\mu$ (homogenous dielectric medium) and $\nabla\varphi_\eta$ (dielectric boundary at the cavity surface):

$$\Delta G_\tau = \tfrac{1}{2}\int_V \left[\left(\varepsilon_\mu - \varepsilon\right)\nabla\varphi_\mu \nabla\varphi_\eta\right] dV \,. \qquad \textbf{(3)}$$

V is the volume of the substituted dielectric body. ΔG_e is a sum of two parts: (I.) Transfer energy of a medium ε_μ into the vacuum outside the molecular cavity and (II.) Transfer energy of a medium ε_v into a system of fixed charges embedded in a homogenous medium with permittivity ε_μ. Following the continuum model solute molecules are approximated by fixed charge distributions embedded in a dielectric cavity with permittivity ε_μ (fig. 1).

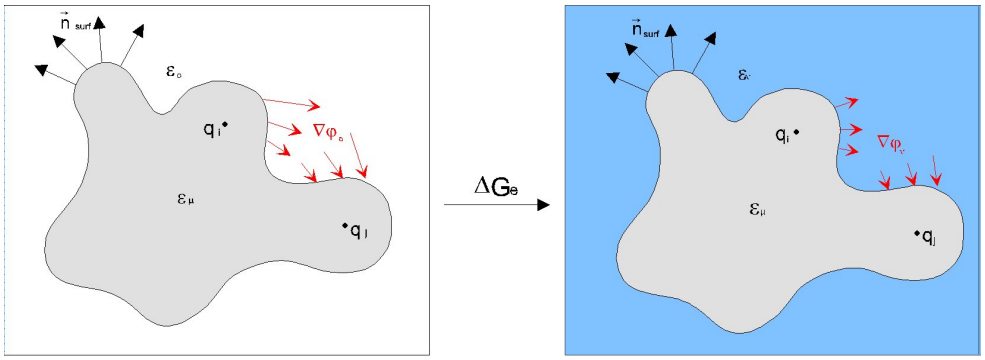

Figure 1: A dielectric continuum representation of a molecular scenario $q_i...q_j$, ε_μ, which is transferred from vacuum, ε_o (left hand side) to a continuous solvent, ε_v. The dielectric solvent weakens the normal components $\vec{n}_{surf} \cdot \nabla\varphi$ of the electrostatic field and thus the total field (red arrows) at the molecular surface.

Using equation (3), the solvation energy of a solute molecule is given by integration over the volume occupied by the solvent:

$$\Delta G_e = \Delta G_{\tau^I - \tau^{II}} = \tfrac{1}{2}\int_V \left[(\varepsilon_\mu - \varepsilon_v)\nabla\varphi_\mu\nabla\varphi_v + (1-\varepsilon_\mu)\nabla\varphi_\mu\nabla\varphi_0\right]dV \qquad \textbf{(4)}$$

where

$$\nabla\varphi_1\nabla\varphi_2 = \nabla(\varphi_1\nabla\varphi_2) + \varphi_1\nabla^2\varphi_2 \qquad \textbf{(5)}$$

The indices μ, ν, and 0 represent solute, solvent, and vacuum, respectively. If the ionic strength of the solvent is negligible, Laplace´s equation $\nabla^2\varphi = 0$ is valid outside of the cavity, where no charges are present:

$$\begin{aligned}\Delta G_e = \tfrac{1}{2}\int_V \big[&\varepsilon_\mu\nabla(\varphi_v\nabla\varphi_\mu) - \varepsilon_v\nabla(\varphi_\mu\nabla\varphi_v) \\ &+ \nabla(\varphi_\mu\nabla\varphi_0) - \varepsilon_\mu\nabla(\varphi_0\nabla\varphi_\mu)\big]dV\end{aligned} \qquad \textbf{(6)}$$

By application of the Gauss-Theorem

$$\int_V \nabla(\varphi_1 \nabla \varphi_2) dV = \oint_S (\varphi_1 \nabla \varphi_2 \cdot \vec{n}_{surf}) \, dS \tag{7}$$

equation (6) is transformed into a surface-integral over the closed surfaces S_{tot} of the solvent along the cavity S_c and at the outer boundary of the solvent S_b.

$$\Delta G_e = \tfrac{1}{2} \oint_{S_{tot}} \{ [(\varepsilon_\mu \varphi_\nu - \varepsilon_\mu \varphi_0) \nabla \varphi_\mu - \varepsilon_\nu \varphi_\mu \nabla \varphi_\nu + \varphi_\mu \nabla \varphi_0] \cdot \vec{n}_{surf} \} dS$$
$$\text{with } S_{tot} = S_c + S_b \tag{8}$$

The terms in rectangular brackets correspond to a scalar energy density functional, which is only nonzero onto the boundaries of the solvent. Since

$$\lim_{r_S \to \infty} \oint_{S_b} (\varphi_1 \nabla \varphi_2 \cdot \vec{n}_{surf}) \, dS = 0, \tag{9}$$

for very dilute solutions ΔG_e is mainly a function of the cavity surface within the restrictions of this model.

3.5 Nonpolar solutes

Experimental measurements related to the hydrophobic effect are based either on the analysis of free energies, enthalpies, and entropies of solvation or the corresponding thermodynamic changes of these quantities for the transfer process from a nonaqueous solvent to water.[6,31,36,37,84-86] For several nonpolar compounds thermodynamic measurements of the hydrophobic hydration at $T = 289$ K result in a positive free energy due to a significant decrease of entropy during the solvation process in water, i.e. one finds $\Delta G_{hyd} > 0$, $\Delta H_{hyd} < 0$ (but small), and $T\Delta S_{hyd} < 0$. For the transfer of nonpolar molecules from an organic solvent to water experimental results show that the enthalpy change contributes dominantly to the total free energy change, i.e. $\Delta G_{hyd} > 0$, $\Delta H_{hyd} > 0$, and $T\Delta S_{hyd} < 0$ (but small).

Abraham et al.[36,37,85,86] constructed a general set of parameters for the solution of gaseous non-electrolytes in various solvents in order to reproduce free energies,

enthalpies, and entropies by the use of simple linear relationships. They used parameters closely correlated to the volume of the solute and to the molecular surface. A correlation between thermodynamic solvation data and molecular surfaces was also established by Privalov et al.[31] All these results indicate that the molecular surface concept is reasonable when discussing the thermodynamic data, in particular the entropy, in a quantitative manner. However, these empirical findings cannot serve as a physical basis for a semi-empirical model.

There has been a considerable effort in the last decades in order to establish a theoretical concept, particularly for the entropic part of the problem. We first refer to the old concept of the *eisberg* model proposed by Frank and Evans.[87] In this model the decrease of entropy in the hydrophobic hydration is explained with the occurrence of a clathrate like hydration shell structure, i.e. the intermolecular hydrogen bond network between water molecules in the first hydration shell leads to a significantly lower degree of freedom in comparison with bulk water in this scenario. It is clear that this entropy contribution can be roughly set proportional to the solvent accessible surface (SAS) of the solute molecule. However, it has been demonstrated[88-90] that the iceberg model is too simple to explain the experimental findings. The decrease of entropy measured at hydrophobic hydration can be rather related to an excluded volume effect while structural effects in the hydration shell tend to increase the solubility in water. An interesting attempt to explain thermodynamic data of solvation of apolar compounds in liquid water is the scaled particle theory (SPT) in which all particles are assumed to be spherical with diameters defined such that thermodynamic quantities of solution and pure liquid are reproduced. Pierotti applied this theory to study hydrophobic effects[91-93] and separated the dissolution of a particle into two steps: first, the creation of a cavity in the solvent which has the appropriate size to accommodate the solute molecule and second, the onset of the interactions between the solute and the solvent. The cavity formation has a basis which is dominantly entropic in nature. It was found that the free energy of cavity formation ΔG_c is roughly proportional to the number of solvent molecules in the first solvation shell[94] (fig. 2). For the case where the size of solute is considerably bigger than that of the solvent molecules the solvation number can be set proportional to the total area of the SAS, i.e. the contribution $T\Delta S_c$ can be represented as a surface integral as well.

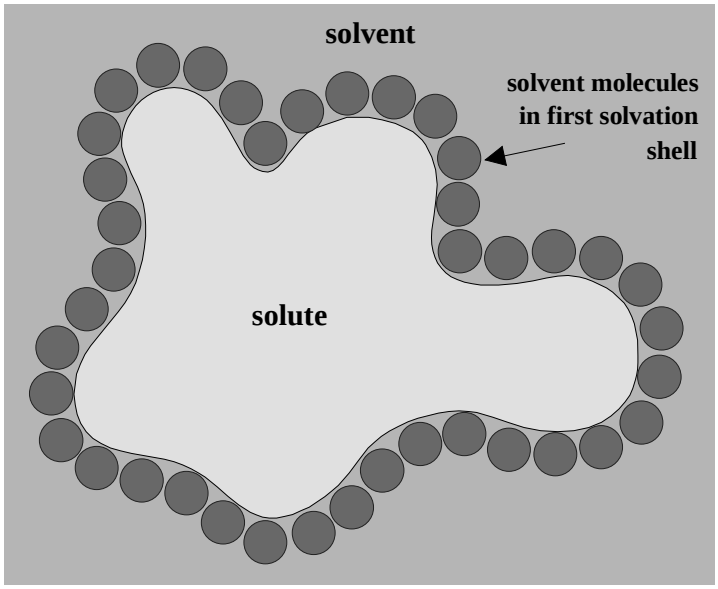

Figure 2: Schematic representation of a solute with its first solvation shell.

In a recent theoretical paper of Ben-Naim et al.[95] on the solvation free energy of a nonpolar hard sphere system a simple relation (10) was established, i.e. in addition to a constant term the free energy difference becomes proportional to the surface and the volume of the solute:

$$\Delta G_{\mathrm{n}} \cong c_0 + c_2 R^2 + c_3 R^3 \qquad \textbf{(10)}$$

Our further considerations are based on this concept. If the diameter of the cavity is significantly larger than the solvent diameter, ΔG_{n} is essentially dominated by the latter terms in eq. (10). We note that the volume V of an arbitrarily shaped object can be written as a surface integral using the divergence theorem[96]

$$V = \int_V dV = \tfrac{1}{3} \oint_S \vec{R} \cdot \vec{n}_{\mathrm{surf}}\, dS\,. \qquad \textbf{(11)}$$

If we neglect the constant term in eq. (10) and apply (11) to the volume dependent term of (10), we obtain a surface integral representation of the free energy of cavity formation

$$\Delta G_n \cong \oint_S F(\vec{R}_S)dS \qquad \textbf{(12)}$$

with a properly chosen function F of the surface points $\vec{R}_S$.

Since purely electrostatic as well as nonpolar terms of the free energy of solvation can be represented by a surface integral, we postulate such a relation to be valid for all intermediate cases:

$$\Delta G_{\text{solv}} = \Delta G_{\text{e}} + \Delta G_{\text{n}} = \int_S \rho(\vec{r}_S)\, dS \qquad \textbf{(13)}$$

wherein $\rho(\mathbf{r}_S)$ is a suitably chosen function of the surface coordinates $\mathbf{r}_S$.

3.6 The empirical molecular free energy surface density (MolFESD) concept

Based on the argumentation presented in the previous section we formulate an empirical model for the transfer free energy $\Delta\Delta G = \Delta G_{\text{transfer}}$ for a substance from aqueous solution to a 1-octanol solution. The experimental data used for the parameterization are taken from partition coefficients P of the substance in a 1-octanol/water system, i.e. strictly speaking we have two liquid phases which are saturated solutions. It has been argued,[6,50,51,97,98] that the octanol phase can be adequately considered as a hydrophobic biophase analogue. This argument will become important in what follows below.

We started from the assumption of Richards et al.[33,42] that the transfer free energy $\Delta G_{\text{transfer}}$ for one mole substance from one solvent to the other can be composed of additive contributions which are related to the solvent accessible surface of fragments of the solute.

$$\Delta G_{\text{transfer}} = -2.303\, RT \log P = \sum_i \Delta G_i \qquad \textbf{(14)}$$

In our approach the $\Delta G_{transfer}$-value does not occur as a sum but as a surface integral over a quantity which is termed *molecular free energy surface density* (MolFESD) given the symbol ρ.

$$\Delta G_{transfer} = \int_S \rho(\vec{r})dS \qquad \textbf{(15)}$$

The quantity $\rho(\mathbf{r})$ measures to what extent a certain surface element contributes to the $\Delta G_{transfer}$-value per unit surface. This notion is similar to the summation of discrete surface elements in the work of Luque et al.[48] The MolFESD value has a physical meaning only on the molecular surface. It corresponds to the surface free energy per unit area in macroscopic liquid/liquid two phase systems or to the surface tension in liquid/gas systems. Insofar our approach is an extrapolation of thermodynamic concepts to a molecular scale.

Similar to eq. (14), the MolFESD can be represented as a superposition of local contributions

$$\rho(\vec{r}_s) = \sum_i \rho_i(\vec{r}_s) \qquad \textbf{(16)}$$

On the basis of this representation the incremental contributions to the total free energy (14) follow as

$$\Delta G_i = \int_S \rho(\vec{r}_S)dS\,. \qquad \textbf{(17)}$$

The summation in (16) is taken over atomic or molecular increments *i*. The incremental contributions ρ_i are chosen in a similar way as the mapping functions considered earlier.[22] Only atomic increments were considered and it was assumed that ρ_i can be adequately modeled by

$$\rho_i(\vec{r}_s) = F_i\,\mu_i(\vec{r}_s) \qquad \textbf{(18)}$$

with an increment specific constant F_i and a membership function μ_i in the sense of fuzzy set theory[99] which measures to what extend a surface point is associated to the increment *i*. We used the membership function which was introduced earlier[100]:

$$\mu_i(\vec{r}_S) = \frac{\exp(-2c_i/\delta_i)+1}{\exp(2(d_i - c_i)/\delta_i)+1} \quad \text{with} \quad d_i = |\vec{r}_S - \vec{r}_i| \tag{19}$$

and a normalization

$$\sum_i \mu(d_i; c_i, \delta_i) = 1 \tag{20}$$

The characteristic proximity parameters c_i and δ_i determine in what way an increment *i* influences the MolFESD-values at a certain surface point.

The membership function μ has a priori no physical meaning. It fulfils two conditions: It is smooth and has finite values for $r < c$ where *c* is a cutoff value which is termed the *proximity distance* of an increment. The value of the parameter *c* should be larger than any van der Waals radius of the increment in the molecule under consideration. For distances $r > c$ the function values of μ rapidly tend towards zero, i.e. the corresponding increment does not significantly contribute to the overall value of the MolFESD. It should be noted here that the individual membership functions implicitly depend on all atomic coordinates of the molecular system as a consequence of the choice of the normalization condition.

The MolFESD function represents a weighted average of all the F_i values of atomic increments *i* for which $d_i < c_i$ is fulfilled, where d_i is the distance from a surface point to the center of the *i*-th increment. All atoms which are further away from the surface point do not contribute significantly.[22,100] The quantities μ_i are related to weighting factors p_i for the contributions of individual atomic free energy surface densities F_i to the local MolFESD-value.

$$\rho(\vec{r}_S) = \sum_i \frac{\mu(d_i; c_i, \delta_i)}{\sum_j \mu(d_j; c_j, \delta_j)} \cdot F_i = \sum_i p_i \cdot F_i \tag{21}$$

The proximity parameters c_i and δ_i for all atomic increments have been determined in the group of the authors.[21] Similar to eq. (21) the quantity

$$S_i = \int_S p_i(\vec{r}_S)dS \qquad \textbf{(22)}$$

can be interpreted as the contribution of the increment *i* to the total molecular surface *S*, and one simply has

$$S = \sum_i S_i \qquad \textbf{(23)}$$

One should note here that although the total surface is given as a sum of increment surfaces S_i the latter are not strictly local properties. A given surface point can only be associated to an increment with a certain weight. There is no sharp borderline between the individual S_i areas in contrast to the treatment of other authors.[42,43,48]

Furthermore, the surface increments do depend on the relative positions of all the other increments, so S_i is explicitly dependent on the molecular conformation. The same atomic increment *i* will, in general, contribute to the total $\Delta G_{\text{transfer}}$-value in a different way, even if the F_i-value of the increment is the same.[100] This fact opens the possibility to include the conformation of a given compound in an empirical parametrization strategy.

Following eqs.(14) and (15) the overall hydrophobicity of a molecule (represented by log *P* data) can be easily calculated. The results are equivalent those which are generated by the MOLCAD program developed in our group. A representative dataset has been recently studied with several popular computational descriptors of hydophobicity.[101] In a multivariate analysis on a database of roughly 160 molecules, similar to other surface-based methods like HINT,[46,47] the total log *P* values derived by MolFESD proved satisfactory.

The MolFESD concept has further been applied[102] in order to estimate the optimal position and orientation of a solute molecule at the interface region of two liquid phases. The argumentation therein was the following: Let ΔG_{I} and ΔG_{II} be the free energies for solving a molecule in solvent I and II, respectively, the transfer free energy $\Delta G_{\text{transfer}}$ is simply given as

$$\Delta G_{\text{transfer}} = \Delta G_{II} - \Delta G_{I} \quad \textbf{(24)}$$

Following the concepts lined out above, the free energy ΔG_i of solvation of a solute molecule in solvent (i) can be described as

$$\Delta G_i = \int_S \rho_i(S)\, dS \quad \textbf{(25)}$$

wherein $\rho_i(S)$ is the free energy surface density for the solute molecule in solvent (i). Comparing equation (24) to equation (25) one has

$$\rho_T(S) = \rho_{II}(S) - \rho_I(S) \quad \textbf{(26)}$$

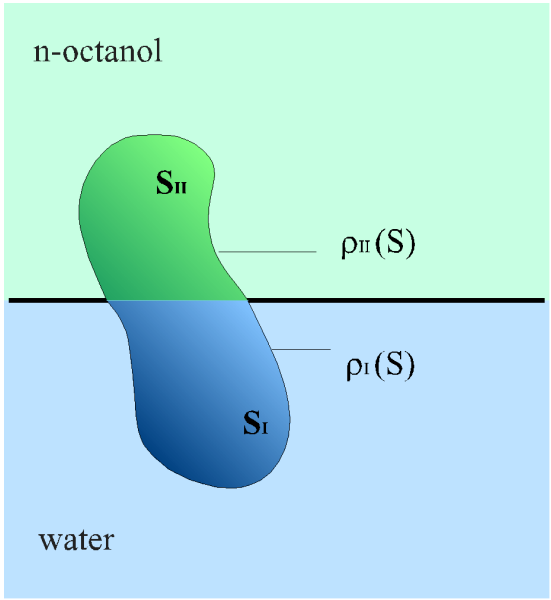

Figure 3: Optimal alignment of a molecule at a phase boundary

The consequent application of the MolFESD concept to a molecule which is brought from the vacuum to a liquid/liquid interface region (fig. 3) leads to

$$\Delta G_{\mathrm{INT}} = \Delta G_{\mathrm{INT}}^{\mathrm{I}} + \Delta G_{\mathrm{INT}}^{\mathrm{II}} \tag{27}$$

$$\text{with} \quad \Delta G_{\mathrm{INT}}^{i} = \int_{S_i} \rho_i(S)\, dS_i \tag{28}$$

Using equation (28) one obtains with $S_{\mathrm{I}} + S_{\mathrm{II}} = S$

$$\begin{aligned} \Delta G_{\mathrm{INT}} &= \int_{S_{\mathrm{I}}} \rho_{\mathrm{I}}(S)\, dS + \int_{S_{\mathrm{II}}} \rho_{\mathrm{I}}(S)\, dS + \int_{S_{\mathrm{II}}} \rho_{\mathrm{T}}(S)\, dS \\ &= \Delta G_{\mathrm{I}} + \int_{S_{\mathrm{II}}} \rho_{\mathrm{T}}(S)\, dS \end{aligned} \tag{29}$$

The first part of this expression is independent from the position and orientation of the molecule in the surface region while the second is. Obviously ΔG_{INT} becomes a minimum – related to a stable placement of the molecule – when the second term in equation (29) becomes minimal. This condition was used in order to calculate the immersion depths of various local anaesthetic in a two liquid hydrophobic/hydrophilic interface.[102] A generalization of this concept is used in the next section for quantitative treatment of an induced fit situation of a highly flexible hydrophobic pocket of a receptor to various sweeteners.

3.7 Quantification of hydrophobic contributions to the sweetness recognition

To our best knowledge the receptor for the recognition of sweeteners is not known up to now. It may happen that there is not only one receptor but several different ones. Therefore in this paper we only will consider substances which are derivatives of sucrose and it seems to be reasonable to assume that all of them are recognized by the sucrose receptor. The current conception of this receptor model dates back to Shallenberger[103] and Kier,[104] who postulated that the sweetness receptor should contain at least three structural components: a proton donor functionality AH, a proton acceptor position B, and a hydrophobic binding site X, all

arranged in a triangle conformation – the sweetness triangle. Lichtenthaler and coworkers[105] postulated on the basis of hydrophobic maps introduced by the group of the present authors[22] that the hydrophobic site should be a rather extended region. A correlation was found between the relative sweetness of several sucrose derivatives and their log *P* data derived by a Crippen-type approach.[106] However, there has been no attempt to quantify hydrophobic contributions to the free energy of ligand-receptor binding.

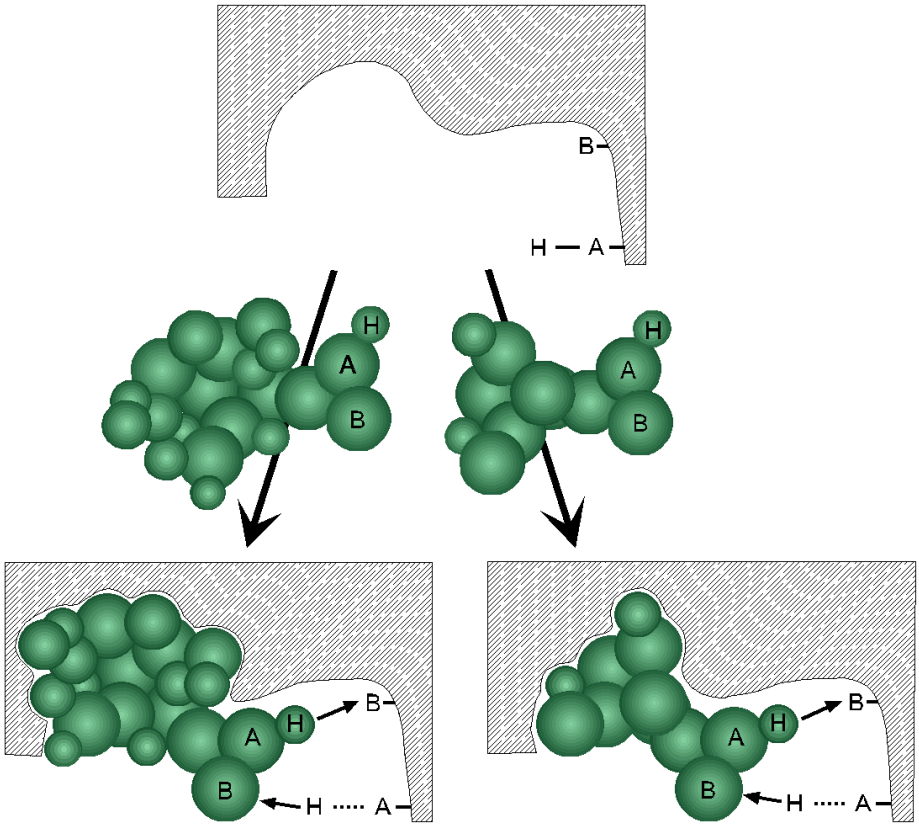

Figure 4: Association of two different molecules to a highly flexible hydrophobic receptor pocket (schematically).

In this work we demonstrate that the MolFESD concept is capable to describe a quantitative local representation of hydrophobicity on the SAS of ligand molecules. In particular we show that the surface integral of MolFESD due to that part of the hydrophobic surface of the ligand explicitly interacting with the receptor correlates well with the free energy of binding. Therefore the method can be used to calculate

the energetic and (implicitly) entropic contribution of a hydrophobic part of a ligand to the total ligand-receptor interaction.

Our model approach has been developed in two steps (model a and b). The first two assumptions for both models can be summarized as follows:

(i) As other authors[107] we suppose that the sweetness is proportional to the equilibrium constant *K* for the association of a sweetener and its receptor.

(ii) We followed the results of Shallenberger,[103] Kier,[104] and Lichtenthaler[105] for the sweetness receptor model. In contrast to these authors, however, in our approach the hydrophobic region is highly flexible, i.e. it can cover a hydrophobic part of the sweetener quantitatively without investing conformational free energy.

(iii) We further adopt that a water saturated 1-octanol solution can be considered as a hydrophobic biophase analogue.

(iiia) In model (a) a highly flexible hydrophobic receptor pocket is emulated as stated above (fig. 4). In this case the hydrophobic contribution to the overall ΔG-value for the association of the molecule to the receptor can be calculated in an analogous manner as has been described in the last section, i.e. one has to minimize

$$\Delta G_{\mathrm{INT}} = \int_{S_{\mathrm{R}}} \rho_{\mathrm{T}}(S)\, dS \qquad \textbf{(30)}$$

where S_{R} is the surface area of hydrophobic interaction between the substrate and the receptor.

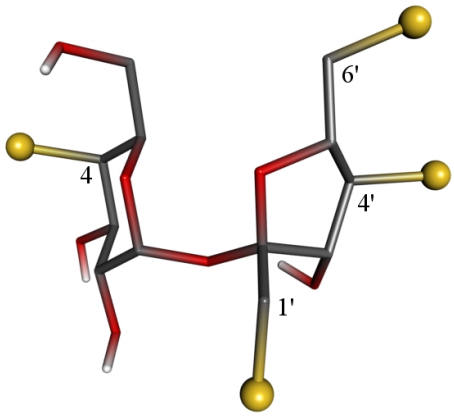

Figure 5: Backbone of the sweetener sucrose – energetic favorable conformation. Substitution positions enhancing sweetness are marked yellow.

Table 1: Relative sweetness and correlation data of sucrose chlorine derivatives calculated from model (a).

	Substitution Positions	Surface Increment, [Å²]	Surface Integral,(log *P*)	ΔG_{INT} [kJ/mol]	Relative Sweetness
1	-	3.160	0.001	0.010	1 [108]
2	1'-Cl	35.423	0.217	-1.239	20 [108]
3	6'-Cl	33.185	0.209	-1.191	20 [108]
4	1'-Cl, 4'-Cl	83.789	0.542	-3.091	30 [109]
5	1'-Cl, 6'-Cl	76.482	0.459	-2.617	80 [110]
6	1'-Cl, 4'-Cl, 6'-Cl	130.135	1.323	-7.549	100 [110]

Before introducing a refinement of model (a), the results will be discussed. The minimization of the hydrophobic surface integral was applied to a series of sucrose and 4-chloro-4-deoxy-*galacto*-sucrose halogen derivatives (table 1, table 2, fig. 5) to exhibit the relation between the hydrophobic surface and the relative sweetness of the selected derivatives. In this case the ΔG-values should only differ by the

hydrophobic contribution ΔG_{INT} which is obviously a minimum value when the integration area in eq. (30) is taken as that one where $\rho_T(S)$ takes a positive value. This is equivalent to the assumption that the receptor will in any case cover all of the hydrophobic part of the molecules under consideration. In table 1 the calculated data based on model (a) and the corresponding sweetness[107] is shown for some chlorine derivatives of sucrose. In table 2 the same information is given for the 4-chloro-4-deoxy-*galacto*-sucrose halogen derivatives.

Table 2: Relative sweetness and correlation data calculated from model (a) of halogen derivatives of 4-chloro-4-deoxy-*galacto*-sucrose

	Substitution Positions	Surface Increment, [Å²]	Surface Integral, (logP)	ΔG_{INT} [kJ/mol]	Relative Sweetness
7	-	25.185	0.099	-0.562	5 [108]
8	6'-Cl	65.902	0.341	-1.946	50 [111]
9	1'-Cl	60.527	0.330	-1.884	120 [111]
10	4'-Cl, 6'-Cl	121.240	1.075	-6.132	160 [111]
11	1'-Cl, 4'-Cl	112.652	0.667	-3.808	220 [111]
12	1'-Cl, 6'-Cl	114.549	0.626	-3.574	600 [110]
13	1'-6'-Cl, 4'-F	156.320	1.181	-6.736	1000 [111]
14	1'-Cl, 4'-Cl, 6'-Cl	169.970	1.531	-8.735	2200 [111]
15	1'-Cl, 6'-Cl, 4'-Br	171.891	1.614	-9.210	3000 [111]
16	1'-6'-Cl, 4'-I	177.928	2.140	-12.211	3500 [111]

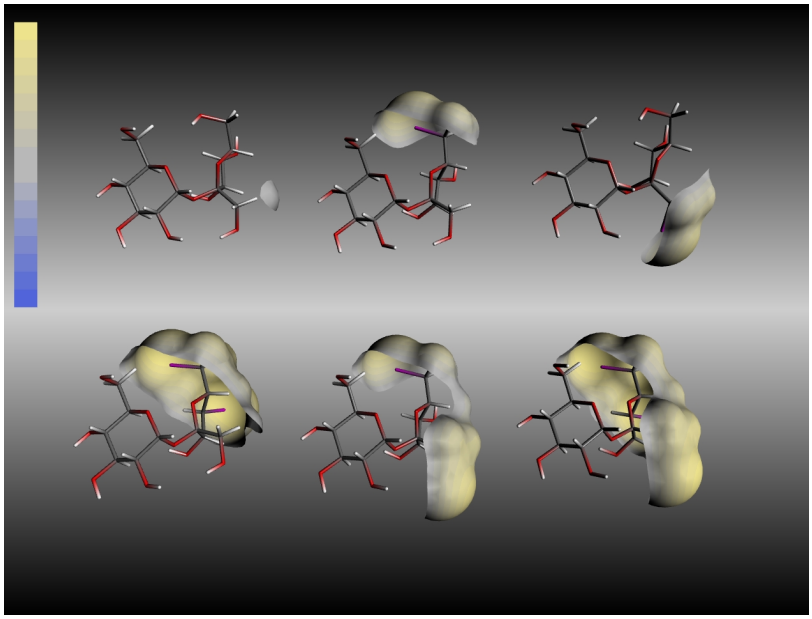

Figure 6: Hydrophobic surface patches with colour coded MolFESD-values (yellow: negative/hydrophobic, blue: positive/hydrophilic) of different chlorine derivatives of sucrose calculated on the basis of model (a). The visualisation was performed with the MOLCAD package.[112]

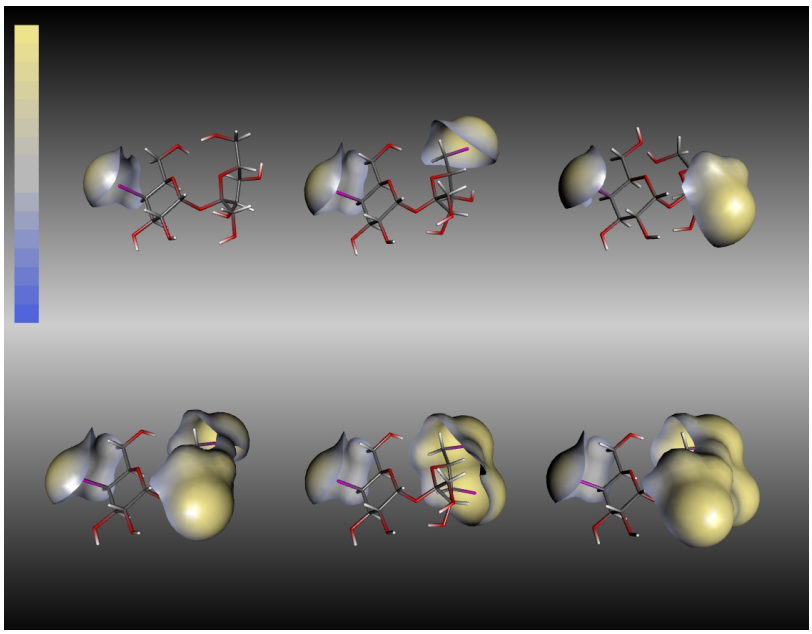

Figure 7: Hydrophobic surface patches with color coded MolFESD-values (as in fig. 6) of different chlorine derivatives of the 4-Chloro-4-deoxy-galacto-sucrose calculated on the basis of model (a).

The natural logarithm of the sweetness data (ln *K*) versus the ΔG_{INT} values corresponding to the hydrophobic surface integral for the sucrose and the 4-chloro-4-deoxy-*galacto*-sucrose derivatives, respectively are drawn in fig. 8. A correlation is found between ln *K* and ΔG_{INT} for most of the experimental data but this correlation is obviously non-linear. In particular the less halogenated molecules show smaller ΔG_{INT} values than those expected from a linear extrapolation of the multiple substituted molecules. This result indicates that the hydrophobic pocket of the receptor is obviously not flexible enough in order to fit only to the hydrophobic parts of the MolFESD, but covers also a certain hydrophilic area where the MolFESD becomes negative.

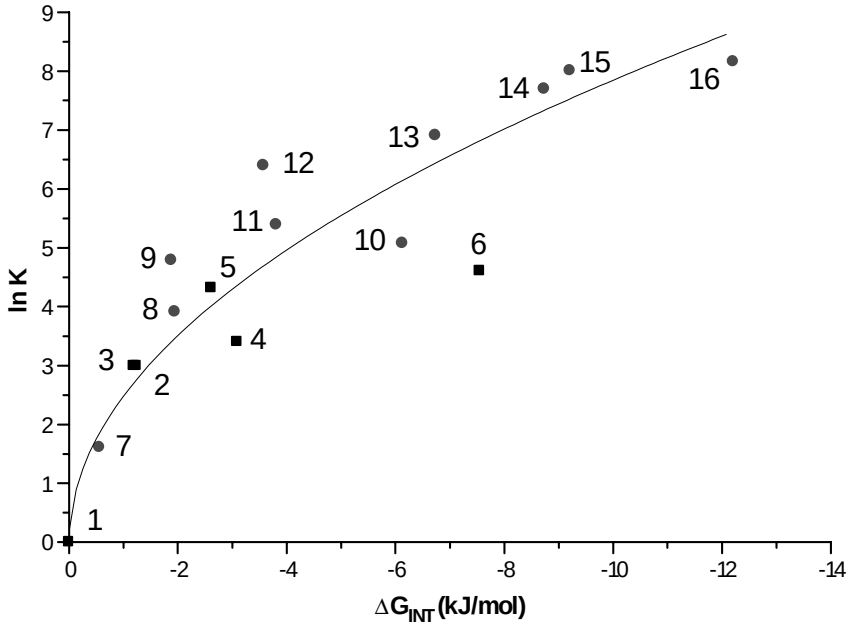

Figure 8: Relative sweetness ln K as a function of ΔG_{INT} (as calculated on the basis of model (a) for halogen derivatives of sucrose (squares) and 4-chloro-4-deoxy-*galacto*-sucrose (circles). The numbers correspond to tables 1 and 2.

A refined model (b) was formulated, wherein it was assumed that

(iiib) for all molecules under consideration most of the furanoid ring is covered by a flexible hydrophobic receptor pocket. The integration over the MolFESD

has to be done along the surface of the furanoid ring and all remaining hydrophobic parts of the surface.

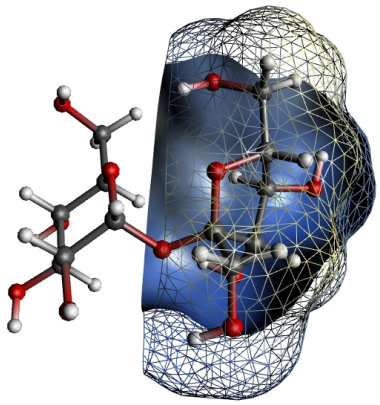

Figure 9: The integration area covering the fructose portion of sucrose is given by the intersection of a plane perpendicular to the C_1-$C_{2'}$ interconnection.

Based on this assumption the borderline of the integration area covering the fructose portion was defined as the intersection of a plane perpendicular to the C_1-$C_{2'}$ interconnection (see fig. 9). A slight shift of this plane backwards and forwards resulted in a collective ΔG_{INT} shift but did not influence the correlation.

Accounting for the hydrophobic interaction of the galactose portion with the receptor, the ΔG_{INT} values of all 4-chloro-4-deoxy-*galacto*-sucrose derivatives were shifted by an additive constant ΔG_{shift}= - 6.8 kJ/mol corresponding to the difference of the calculated log *P* values between sucrose and 4-chloro-4-deoxy-*galacto*-sucrose.

Based on this argumentation the ΔG_{INT} values for the same series of sucrose and the 4-chloro-4-deoxy-*galacto*-sucrose halogen derivatives were calculated (tables 3 and 4).

Table 3: Relative sweetness and correlation data from model (b) of chlorine derivatives of sucrose

	Substitution Positions	Surface Increment, [Å²]	Surface Integral, (logP)	ΔG_{INT} [kJ/mol]	Relative Sweetness
1	-	139.644	-1.730	0	1 [108]
2	1'-Cl	163.243	-1.257	-2.7	20 [108]
3	6'-Cl	157.286	-0.998	-4.176	20 [108]
4	1'-Cl, 4'-Cl	170.018	-0.081	-9.407	30 [109]
5	1'-Cl, 6'-Cl	167.468	-0.229	-8.562	80 [110]
6	1'-Cl, 4'-Cl, 6'-Cl	179.656	1.060	-15.918	100 [111]

Table 4: Relative sweetness and correlation data from model (b) of halogen derivatives of 4-chloro-4-deoxy-*galacto*-sucrose

	Substitution Positions	Surface Increment, [Å²]	Surface Integral, (logP)	ΔG_{INT} [kJ/mol]	Relative Sweetness
7	-	146.534	-1.793	-6.439	5 [108]
8	6'-Cl	157.780	-0.927	-11.379	50 [111]
9	1'-Cl	154.785	-0.968	-11.147	120 [111]
10	4'-Cl, 6'-Cl	164.871	0.318	-18.487	160 [111]
11	1'-Cl, 4'-Cl	166.641	0.007	-16.711	220 [111]
12	1'-Cl, 6'-Cl	168.201	-0.146	-15.836	600 [110]
13	1'-6'-Cl, 4'-F	163.861	0.817	-21.334	1000 [111]
14	1'-Cl, 4'-Cl, 6'-Cl	174.595	1.182	-23.414	2200 [111]
15	1'-Cl, 6'-Cl, 4'-Br	178.468	1.258	-23.851	3000 [111]
16	1'-6'-Cl, 4'-I	180.131	1.828	-27.101	3500 [111]

Plotting the natural logarithm of the sweetness data (ln *K*) versus the ΔG_{INT} values of the hydrophobic surface integral, we obtain a linear relationship with a regression line showing a correlation coefficient of r=0.94 and σ=0.81 for the sucrose and the 4-chloro-4-deoxy-*galacto*-sucrose derivatives (fig. 10).

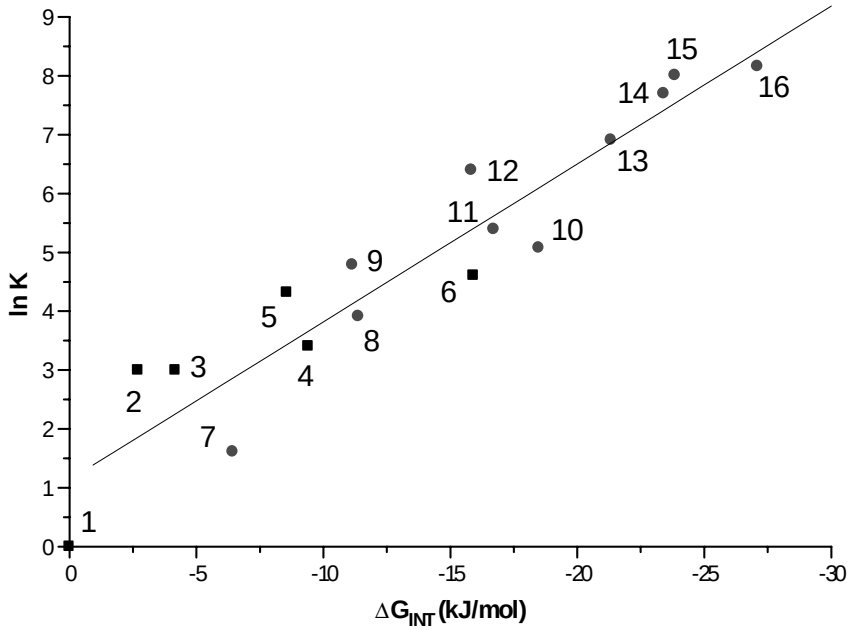

Figure 10: Relative sweetness ln *K* as a function of ΔG_{INT} (as calculated on the basis of model (b) for halogen derivatives of sucrose (squares) and 4-chloro-4-deoxy-*galacto*-sucrose (circles): r=0.94 and σ=0.81.). The numbers correspond to tables 3 and 4.

Our results clearly demonstrate that the MolFESD approach can be adequately used to localize hydrophobicity on the molecular surface. It allows a straightforward interpretation of individual contributions of local surface patches to the overall free energy value. It further renders to quantify the hydrophobic contribution to the total free energy of association of different halogen substituted sucrose derivatives to a potential receptor even if the receptor has not been identified yet.

3.8 Summary and Conclusions

It has been demonstrated in this paper that the free energy of solvation of molecules in polarizable as well as apolar solvents can be represented as a surface integral over a suitably chosen function $\rho(S)$ – termed molecular free energy density, MolFESD – along the solvent accessible surface (SAS) of the molecule. The

introduction of the MolFESD concept is justified by an electrostatic model describing the interaction of a molecular charge distribution with a continuous solvent separated by the SAS. This interaction can be transformed to a surface integral representation using the Gauss theorem. For nonpolar solutes the representation of the free energy of solvation is based on statistical arguments. This quantity becomes a surface extensive property if the ratio of solvent size to solute size tends towards zero. As a consequence of the result in the limiting cases we proposed that the MolFESD concept can generate reasonable estimates of solvation free energies or transfer free energies between two liquid solvents in all intermediate cases.

The MolFESD approach offers a localized picture of molecular hydrophobicity on the SAS with the benefit of a quantitative interpretation. This is accomplished by the fact that the MolFESD function can be composed of atom based contributions wherein membership functions control the individual surface dues. The parameterization has been established on the basis of numerous experimental data for the partition coefficient in an 1-octanol/water system.

The MolFESD concept has been applied to the calculation of the hydrophobic contribution ΔG_{INT} to the overall ΔG-value for the association of several halogen derivatives of sucrose and 4-chloro-4-deoxy-*galacto*-sucrose to the (still unknown) sweetness receptor. In our first approach the original receptor model of Shallenberger and Kier was modified by means of a highly flexible receptor pocket which is able to cover all of the hydrophobic part of the ligands. Taking a water-saturated 1-octanol solution as a hydrophobic biophase analogue, we have calculated ΔG_{INT}-values. The comparison between these values and the experimentally determined sweetness of the compounds exhibited a clear but non-linear correlation.

A second, refined model was formulated, wherein it was assumed that most of the fructose portion is covered by a flexible hydrophobic receptor pocket. In order to estimate the ΔG_{INT}-values the integration over the MolFESD was done along the surface of the furanoid ring whereas other hydrophobic parts of the surface were taken as additive contributions. Having done the same comparison as stated above we found a linear correlation of experimental sweetness and ΔG_{INT}, proving the receptor model assumptions to be convenient. Therefore our model should be

capable to predict relative sweetness for new compounds which only differ in modifications in the hydrophobic part of sucrose.

It is clear that the model approach presented here has its limitations if the microscopic structure of the solvation shell has to be taken into account and explicit calculations of conformational, energetic and entropic terms have to be performed. We are nevertheless optimistic that the MolFESD concept can be very effectively used to generate new physical properties (like partial ΔG_{INT} values) which can be efficiently used in QSAR studies for the prediction of active compounds.

3.9 References

1) Blokzijl,W. and Engberts, J.B.F.N., Angew. Chem. Int. Ed. Engl. 32**,** (1993) 1545.
2) Tanford, C., The Hydrophobic Effect: Formation of Micelles and Biological Membranes. Wiley, New York, 1973.
3) Fujita, T., Iwasa, J. and Hansch, C., J. Am. Chem. Soc. 86, (1964) 5175.
4) Nys, G. C. and Rekker, R. F., Chim. Therap. 8, (1973) 521.
5) Rekker, R. F., The Hydrophobic Fragmental Constants, Elsevier, New York (1977).
6) Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York (1979).
7) Broto, P., Moreau, G. and Vandycke, C., Eur. J. Med. Chem. - Chim. Ther. 19, (1984) 71.
8) Ghose, A. K. and Crippen, G. M., J. Comput. Chem. 7, 4, (1986) 565.
9) Ghose, A. K., Pritchett, A. and Crippen, G. M., J. Comput. Chem. 9, 1, (1988) 80.
10) Viswanadhan, V. N., Ghose, A. K., Revankar, G. R. and Robins, R. K., J. Chem. Inf. Comput. Sci. 29, (1989) 163.
11) Ghose, A.K., Viswanadhan, V.N., Wendoloski, J.J., J. Phys. Chem. A, 102, (1998) 3762.
12) Buchwald, P., Bodor, N., Curr. Med. Chem. 5, (1998), 353.
13) Kantola, A., Villar, H.O. and Loew, H., J. Comput. Chem. 12, (1991) 681.
14) Alkorta, I. and Villar, H.O., Intern. J. Quantum Chem. 44, (1992) 203.
15) Abraham, D.J. and Leo, A.J., Proteins Struct. Funct. Gen., (1987) 130.
16) Kellogg, G.E., Joshi, G.S. and Abraham,D.J., Med. Chem. Res. 1, (1992) 444.
17) Kellogg, G.E. and Abraham, D.J., J. Mol. Graphics 10, (1992) 212.
18) Kellogg, G.E., Semus, S.F. and Abraham, D.J., J. Comput. Aided Mol. Design 5, (1991) 545.
19) Abraham, D.J. and Kellogg, G.E., in "3D QSAR in Drug Design: Theory, Methods and Applications", Kubinyi, H. Ed., Escom, Leiden, (1993) 506.
20) Wildman, S.A. and Crippen, G.M., J. Chem. Inf. Comput. Sci. 39, (1999) 868.
21) Jäger R. and Brickmann J., Empirical quantification of hydrophobicity with the MolFESD strategy, 1999, in preparation.
22) Heiden, W., Moeckel, G. and Brickmann, J., J. Comput. Aided Mol. Des. 7, (1993) 503.
23) Connolly, M., Science 221, (1983) 709.
24) Audry, E., Dubost, J.P., Colleter, J.C. and Dallet, P., Eur. J. Med. Chem. Chim. Ther. 21, 1, (1986) 71.
25) Fauchère, J.-L., Quarendon, P. and Kaetterer, L., J. Mol. Graphics 6, 4, (1988) 203.
26) Brasseur, R., J. Biol. Chem. 266, (1991) 16120.
27) Furet, P., Sele, A. and Cohen, N.C., J. Mol. Graphics 6, 4 (1988) 182.

28) Gaillard, P., Carrupt, P.A., Testa, B. and Boudon, A., J. Comput. Aided Mol. Design 8, (1994) 83.
29) Carrupt, P.A., Gaillard, P., Billois, F., Weber, P., Testa, B., Meyer, C. and Pérez, S., Lipophilicity in Drug Research, Pliska, V., Testa, B., van de Waterbeemd, H. (ed.),. VCH Publishers, Weinheim (1996).
30) Gaillard, P., Carrupt, P.A., Testa, B. and Schambel, P., J. Med. Chem. 39, (1996) 126.
31) Privalov, P.L. and Gill, S.J., Adv. Prot. Chem. 39, (1985) 191.
32) Lee, B. and Richards, F.M., J. Mol. Biol. 55, (1971) 379.
33) Cothia, C., J. Mol. Biol. 105, (1976) 1.
34) Hermann, R.B., Proc. Natl. Acad. Sci. USA 74, (1977) 4144.
35) Tanford, C., The Hydrophobic Effect, Wiley, New York, 1980.
36) Abraham, M.H.; J. Am. Chem. Soc. 104, (1982) 2085.
37) Abraham, M.H., J. Chem. Soc. Faraday Trans. 1, 80, (1984) 153.
38) Eisenberg, D. and McLachlan, A.D., Nature 319, (1986) 199.
39) Radzicka, A. and Wolfenden, R., Biochemistry 27, (1988) 1664.
40) DeYoung, L.R. and Dill, K.A., J. Phys. Chem. 94, (1990) 801.
41) Sharp, K.A., Nicholls, A., Fine, R.F. and Honig, B., Science 252, (1991) 106.
42) Richards, N.G.J., Williams, P.B. and Tute, M., Intern. J. Quantum Chem.: Quantum Biol. Symp. 18, (1991) 299.
43) Richards, N.G.J., Williams, P.B. and Tute, M., Intern. J. Quantum Chem. 44, (1992) 219.
44) Hermann, R.B., J. Comput. Chem. 18, (1997) 115.
45) Lee, B., Proc. Natl. Acad. Sci. USA 88, (1993) 5154.
46) Kellogg, G.E., Semus, S.E. and Abraham, D.J., J. Comput.-Aided Mol. Des. 5, (1991) 545.
47) Abraham, D.J., Kellogg, G.E., Holt, J.M. and Ackers, G.K., J. Mol. Biol. 272, (1997) 613.
48) Luque, F.J., Barril, X. and Orozco, M., J. Comput. Aided Mol. Design 13, (1999) 139, Muňoz, J., Barril, X., Luque, F.J., Gelpí, J.L., and Orozco, M., private communication, to be publ.
49) Sinanoglu, O. , J. Chem. Phys., 75, (1981), 463.
50) DeBolt, S.E. and Kollman, P.A., J. Am. Chem. Soc. 117, (1995) 5316.
51) Hansch, C., Dunn, J., III., J. Pharm. Sci., 61, (1972), 1.
52) Davis, M.E. and McCammon, J.A., Chem. Rev. 90, (1990) 509.
53) Warwicker, J. and Watson, H.C.; J. Mol. Biol. 157, (1982) 671.
54) Gilson, M.K., Sharp, K.A. and Honig, B.; J. Comp. Chem. 9, (1987) 327.
55) Sharp, K., Jean-Charles, A. and Honig, B., J. Phys. Chem. 96, (1992) 3822.
56) Sharp, K.A., Nicholls, A., Fine, R.F. and Honig, B., Science 268, (1995) 1144.
57) Rashin, A.A., J. Phys. Chem. 93, (1989) 4664.
58) Juffer, A.H., Botta, E.F., van Keulen, B.A., van der Plog, A. and Berendsen, H.J.C., J. Comput. Phys. 97, (1991) 144.
59) Zauhar, R.J. and Morgan, R.S., J. Comput. Chem. 9, (1988) 171.
60) You, T.J. and Harvey, S.C., J. Comput. Chem. 14, (1993) 484.

61) Gilson, M. and Honig, B., Proteins 4, (1988) 7.
62) Sitkoff, D., Ben-Tal, N. and Honig, B., J. Phys. Chem. 100, (1996) 2744.
63) Zauhar, R.J. and Morgan, R.S., J. Mol. Biol. 186, (1985) 815.
64) Luty, B.A., Davis, M.E. and McCammon, J.A., J. Comput. Chem. 13, (1992) 768.
65) Still, W.C., Tempczyk, A., Hawley, R. and Hendrickson, T., J. Am. Chem. Soc. 112, (1990) 6127.
66) Cramer, C.J. and Truhlar, D.G., J. Am. Chem. Soc. 113, (1991) 8305.
67) Qiu, D., Shenkin, P.S., Hollinger, F.P. and Still, W.C., J. Phys. Chem. A, 101, (1997) 3005.
68) Giesen, D.J., Storer, J.W., Cramer, C.J. and Truhlar D.G., J. Am. Chem. Soc. 117, (1995) 1057.
69) Giesen, D.J., Cramer, C.J. and Truhlar D.G., J. Phys. Chem. 99, (1995) 7137.
70) Chambers, C.C., Hawkins, G.D., Cramer, C.J. and Truhlar D.G., J. Phys. Chem. 100, (1996) 16385.
71) Cramer, C.J. and Truhlar D.G., J. Comput. Aided Mol. Des. 6, (1992) 629.
72) Giesen, D.J., Chambers, C.C., Cramer, C.J. and Truhlar D.G., J. Phys. Chem. B, 101, (1997) 2061.
73) Hawkins, G.D., Cramer, C.J. and Truhlar D.G., J. Phys. Chem. 100, (1996) 19824.
74) Giesen, D.J., Gu, M.Z., Cramer, C.J. and Truhlar D.G., J. Org. Chem. 61, (1996) 8720.
75) Edinger, S.R., Cortis, C., Shenkin, P.S. and Friesner, R.A., J. Phys. Chem. B, 101, (1997) 1190.
76) Cramer, C.J. and Truhlar, D.G., Continuum Solvation Models: Classical and Quantum Mechanical Implementations. In Reviews in Computational Chemistry: Lipkowitz, K.B., Boyd, D. Ed.: VCH New York 1995: Vol.6.
77) Born, M., Z. Phys. 1, (1920) 45.
78) Kirkwood, J., J. Chem. Phys. 3, (1935) 300.
79) Srebrenik, S., Weinstein, H. and Pauncz, R., Chem. Phys. Lett. 20 (5), (1973) 419.
80) Schaefer, M. and Frömmel, C., J. Mol. Biol. 216, (1990) 1045.
81) Schaefer, M. and Karplus, M., J. Phys. Chem. 100, (1996) 1578.
82) Smythe, W.R., Static and Dynamic Electricity, McGraw-Hill, New York 1967.
83) Fröhlich, H., Theory of Electric Polarization, Claredon Press, Oxford 1958.
84) Wilhelm, E., Battino, R. and Wilcock, R.J., Chem. Rev. 77, (1977) 219.
85) Abraham, M.H., Grellier, P.L. and McGill, R.A., J. Chem. Soc. Perkin Trans. 2, (1988) 339.
86) Abraham, M.H. and Nasehzadeh, A., J. Chem. Soc. Faraday Trans. 1, 77, (1981) 321.
87) Frank, H.S. and Evans, M.W.; J. Chem. Phys. 13, (1945) 507.
88) Lee, B., Meth. Enzymol. 259, (1995) 555.
89) Lee, B., Biopolymers 31, (1991) 993.
90) Lee, B., Biopolymers 24, (1985) 813.

91) Pierotti, R.A., Chem. Rev. 76, (1976) 717.
92) Pierotti, R.A., J. Phys. Chem. 69, (1965) 281.
93) Pierotti, R.A., J. Phys. Chem. 67, (1963) 1840.
94) Reiss, H., Adv. Chem. Phys. 9, (1966) 1.
95) Ben-Naim, A. and Lovett, R., J. Phys. Chem. B 101, (1997) 10535.
96) Jackson, J.D., Classical Electrodynamics, Wiley, New York, 1975.
97) Best, S.A:, Merz, K.M. and Reynolds, C.H., J. Phys. Chem. B, 103, (1999) 714.
98) Privalov, P.L., Gill, S.J. and Murphy, K.P., Science 250, (1990) 297.
99) Zimmermann, H.J., Fuzzy Set Theory and its Applications, Kluwer, Boston, 1991.
100) Pixner, P., Heiden, W., Merx, H., Möller, A., Moeckel, G. and Brickmann, J., J. Chem. Inf. Comput. Sci. 34, (1994) 1309.
101) Mannhold, R., Cruciani, W., Dross, K. and Rekker, R., J. Comput.-Aided Mol. Des. 12, (1998) 573.
102) Jäger, R., Segal, A., Flonda, M.L. and Brickmann, J., in preparation.
103) Shallenberger, R.S. and Acree, T.E., Nature 216, (1967) 480.
104) Kier, L.B., J. Pharm. Sci. 61, (1972) 1394.
105) Immel, S., Kreis, U. and Lichtenthaler, F.W., Starch 43, 4, (1991) 121.
106) Immel, S., Ph. D. thesis, Darmstadt University of Technology (1995).
107) Lee, C.K., Adv. Carboh. Chem. Biochem. 45, (1987) 199.
108) Hough, L. and Khan, R., Trends Biol. Sci. 3, (1978) 61.
109) Jenner, M.R., Jackson, G.,Lee, C.K:, and Khan, R.A. (Tate & Lyle PLC), UK Pat. GB2104063 (1983).
110) Jackson, G., Jenner, M.R. and Khan, R.A. (Tate & Lyle PLC), US Pat. US4473546 (1984).
111) Jackson, G., Jenner, M.R., Khan, R.A: Lee, C.K., Mufti, K.S., Patel, G.D. and Rathbone, E.B. (Tate & Lyle PLC) Eur. Pat. EP0073093 (1983).
112) Knoblauch, M., Waldherr-Teschner, M., MOLCAD. Neue Entwicklungen von Molecular-Modeling-Software für Superworkstations. In: Gauglitz, G. (ed.), Software-Entwicklungen in der Chemie 3, Springer Verlag Berlin (1989).

4 Fast Prediction of Hydration Free Energies from Molecular Interaction Fields

4.1 Abstract

A novel empirical model is presented that allows the fast computation of hydration free energies with high accuracy. The linear model is based upon the separation of the free energy of hydration into a cavity and an interaction term. The cavity contribution is modeled as a linear combination of molecular volume and surface terms. The interaction part is derived from the statistical three-dimensional free energy density and is modeled approximately as a molecular interaction field using the program GRID. A compression scheme is employed to represent this three-dimensional information on the molecular surface by means of a linear combination of surface functions. A set of 81 small organic molecules with known experimental hydration free energies is used to determine the coefficients of the linear model by least squares regression. The fit is statistically significant yielding a correlation coefficient of 0.99, a root mean square error of 0.27 kcal/mol for the 81 molecules belonging to the training set, and 0.63 kcal/mol for a small test set of 10 molecules.

4.2 Introduction

Fast and accurate prediction of solvation free energies is the focus of various modern research fields, such as drug design, drug disposition, and many molecular recognition phenomena. Several computational approaches with different levels of complexity for calculating solvation free energies have been established over the past several decades. These approaches comprise methods such as molecular dynamics or Monte-Carlo simulation,[1-5] mixed quantum/molecular mechanics techniques,[6] and the broad field of implicit (continuum) solvent models that use either classical, quantum mechanical, or hybrid methods.[7-10] The quantum-mechanical self-consistent reaction field method[11] was applied by Luque et al.[12] in an attempt to devise a fractional description of the free energy of solvation based on the solvent exposed surface of the solute. The benefit of this approach is the insight one gains regarding

the influence of structural properties on the solvation process. Corrections terms required in many fractional treatments of the transfer free energy[13] now become unnecessary.

On the other hand, several methods based on empirical models exist.[14,15] No et al.[14] have used the concept of a hydration free energy density modeled as a linear combination of physical molecular properties such as atomic charge, atomic polarizability, and a dispersion energy coefficient to predict successfully the free energy of hydration. Viswanadhan and coworkers[15] find a higher predictive capability from a group contribution approach compared to a quantum-mechanical solvation model.

In this work we establish a linear model based on a condensed surface representation of hydration free energy density by combining both the concept of a hydration free energy density and the group contribution method. As an intermediate step toward this model description we generate a molecular interaction field to approximate the interaction part of the hydration free energy density. Such molecular interaction fields play an essential part in many 3D-QSAR[13,16,17] and CoMFA[18] applications. One of the most popular approaches to generate such molecular fields with respect to steric and electrostatic interactions is the program GRID by Goodford.[19] Its concept is based on the definition of a variety of molecular fragments, termed *probes*, resembling small organic functional groups, whose interaction potential is computed for all points of a grid that encloses the target structure under investigation. Characteristic binding features of the ligand can be deduced combining a GRID analysis with certain statistical methods like PLS or PCA.[20] In addition, important variables might be identified by the more advanced GOLPE[21] algorithm.

A more recent development based on the GRID analysis is implemented in the program *VolSurf*.[22] This program intends to derive two-dimensional descriptors from 3D molecular maps. An application of the *VolSurf* algorithm has recently been published by Alifrangis et al.[23] In addition to these CoMFA and 3D-QSAR applications, the ability of the GRID program to identify attractive interaction regions of a possible ligand and a receptor structure also qualifies it as a useful tool for *de novo* design and molecular docking strategies.[24]

This article presents a parameterization strategy that combines local statistical and global thermodynamic properties of the molecule. An optimal balance between

these reference properties is found, yielding an empirical model with strong predictive power. Certain preliminary investigations together with a detailed description of the model itself are presented. After successful parameterization, the resulting set of parameters allows for the rapid calculation of hydration free energies with the input of the three-dimensional structure of the compound.

4.3 Statistical Basis: Three-dimensional free energy of hydration density

The thermodynamics of solvation processes in terms of enthalpic and entropic contributions is described by the Gibbs or Helmholtz free energy of solvation. In what follows, we make the distinction between Gibbs and Helmholtz free energies of solvation only where explicitly needed. In the absence of an external field, i.e. for a uniform fluid, the chemical potential or free energy of solvation of a particle at infinite dilution is related to the three-dimensional free energy density (3D-FED), ρ_{solv}, as

$$\Delta G_{\mathrm{solv}} = \int_V \rho_{\mathrm{solv}}(\mathbf{r})\, d\mathbf{r} \qquad \textbf{(1)}$$

where V is the volume and $\mathbf{r}$ is a spatial vector. This relation has its origin in early work on statistical density functional theory.[25] For a simple liquid with pair potentials $V(\lambda)$ and density ρ, we define

$$\rho_{\mathrm{solv}}(\mathbf{r}) = \rho \int_0^1 g(\lambda, \mathbf{r}) \left(\frac{\partial V(\lambda, \mathbf{r})}{\partial \lambda} \right) d\lambda \qquad \textbf{(2)}$$

where g is the pair distribution function, λ is a coupling parameter governing $V(\lambda=0) = 0$, and $V(\lambda=1) = V$ (extension to multi-component and molecular fluids is straightforward). Inserting this expression into eq. 1, we recover the usual thermodynamic integration formula[1]

$$\Delta G_{\text{solv}} = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \,. \qquad \textbf{(3)}$$

4.4 Model I: The GRID approximation to the 3D-FED

In principle, the 3D-FED as specified in eq. 2 is accessible, e.g. through molecular simulation or integral equation theory. Although it contains all necessary thermodynamic information, evaluating this exact density is computationally demanding. In an alternative approach to simplify the computation, the solvation process can be understood in terms of two successive steps, namely (a) the creation of a cavity according to the size of the solute molecule and (b) the insertion of the solute by switching on interactions with its environment ("charging"). This leads to a subdivision of the free energy of solvation in terms of two parts:

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{int}} \qquad \textbf{(4)}$$

where the subscripts cav/int denote cavity and interaction contribution, respectively. Most of the difference between Gibbs and Helmholtz free energies, the pressure volume work, is contained in the cavitation term, the interaction part is therefore almost independent of the chosen ensemble.

As a consequence of such a partitioning of the 3D-FED, one must carefully divide the total volume V_{total} into the two regions pertaining to the cavity and to the interaction region, respectively. Choosing the solvent accessible surface[26] as a reasonable border separating cavity and interaction regions the volume integral of the free energy density becomes, in analogy to eq. 4

$$\int_{V_{\text{total}}} \rho_{\text{solv}}(\mathbf{r})\, d\mathbf{r} = \int_{V_{\text{cav}}} \rho_{\text{cav}}(\mathbf{r})\, d\mathbf{r} + \int_{V_{\text{int}}} \rho_{\text{int}}(\mathbf{r})\, d\mathbf{r} \,. \qquad \textbf{(5)}$$

In the following, the term *molecular surface* is taken as the solvent accessible surface described by Conolly.[26]

In our approach, the interaction part of the 3D-FED, ρ_{int}, is approximated by the molecular interaction field generated from GRID. This program is well suited for

two reasons: the fast generation of interaction fields, i.e. 40 CPU seconds are needed on a MIPS R5000 processor (150 MHz) for a molecule with 145 atoms, and the reliable internal force field that contains both enthalpic and entropic contributions. For the modeling of the cavity contribution a linear combination of molecular surface and volume terms is applied.[14,27] Therefore, ΔG_{cav} and ΔG_{int} become

$$\Delta G_{\mathrm{cav}} = \int_{V_{\mathrm{cav}}} \rho_{\mathrm{cav}}(\mathbf{r}) d\mathbf{r} \approx \beta S + \gamma V + \mathrm{const.} \qquad \textbf{(6)}$$

$$\Delta G_{\mathrm{int}} = \int_{V_{\mathrm{int}}} \rho_{\mathrm{int}}(\mathbf{r}) d\mathbf{r} \approx \alpha \sum_i \varepsilon_i \qquad \textbf{(7)}$$

where S is the molecular surface, V is the volume enclosed by this surface, ε_i is the energy value of grid point i, α, β, γ are the parameters, and *const*. is a residual contribution due to a small cavity. We now focus on the *hydration* free energy, ΔG_{hyd}. In this case, the radius of the solvent is much smaller than the radius of the solute, so the constant term in the linear combination is neglected. The total linear model therefore reads

$$\Delta G_{\mathrm{hyd}} = \alpha \sum_i \varepsilon_i + \beta S + \gamma V \,. \qquad \textbf{(8)}$$

We used the water probe supplied by GRID for a set of 81 molecules with known experimental hydration free energies taken from No et al.[14] The total energy for a grid point i, ε_i, is calculated by GRID from the sum of Lennard-Jones, hydrogen bonding, and Coulomb potential functions as well as an additional entropy term that accounts for the hydrophobic hydration effects. Although the entropy estimation is rather simple in nature, it allows us to interpret the resulting interaction field, divided by the volume element the grid point occupies, as a free energy density.

The clearance of the grid was set to 4Å with a spacing of 0.5Å for the grid points. This gives an average number of approximately 17,000 grid points per molecule. The summation according to eq. 7 was carried out for each molecule over all grid points i with a *negative* energy value. The molecular surface and volume were computed[26] for each molecule and the parameters α, β, and γ were then

determined by linear least squares regression. The fit is significant (F = 285.16) yielding a correlation coefficient R^2 = 0.92 with a mean error of σ = 1.15 kcal/mol. Figure 1 shows the predicted response values plotted against the experimental free energies of hydration.

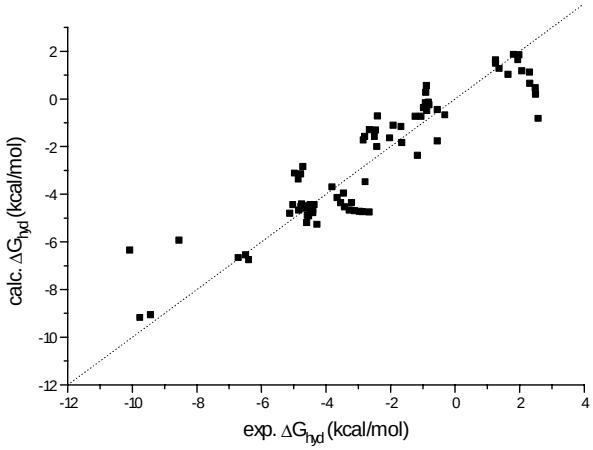

Figure 1 Calculated free energies of hydration vs. experimental values, as derived from the simple GRID model I (eq. 8).

After successfully applying the molecular interaction field to approximate the interaction part of the hydration free energy, further optimization of the model was pursued in two steps: First, the grid information can be compressed to a surface-based representation thereby improving computational speed. Then, the parameters of the compressed form can be optimized with respect to both local grid data and global thermodynamic information. The first aspect is described in the next part of this paper, the second is presented in the section “Final Model”.

4.5 Model II: Expansion of the 3D-FED in surface-based functions

The GRID-based interaction part of the hydration free energy is used as input in a compression scheme allowing for even faster computation retaining as much information as possible. Starting from the molecular interaction field, the 3D-FED is represented as a linear combination of a set of basis functions, f_i, located on distinct patches, i, on the molecular surface. These functions are chosen in a manner that

combines an exponential distance dependence with a spherical harmonics expansion, i.e.

$$f_j(\mathbf{r}-\mathbf{r}_i) = P^j(\cos\vartheta)\, e^{-k|\mathbf{r}-\mathbf{r}_i|} \quad \textbf{(9)}$$

with P^j being the Legendre polynomial of degree j

$$P^j(\cos\vartheta) = \frac{1}{2^j\, j!}\frac{d^j}{d\cos\vartheta^j}(\cos^2\vartheta - 1)^j \quad \textbf{(10)}$$

where $\mathbf{r}_i$ is a reference point on patch i. ϑ denotes the angle between the normal vector of the patch reference point n_i and $\mathbf{r}$-$\mathbf{r}_i$. The interaction part of the 3D-FED is then written as

$$\rho_{\text{int}}(\mathbf{r}) = \alpha\frac{\varepsilon(\mathbf{r})}{\Delta V} \approx \alpha\sum_i\sum_j a_{ij}\, f_j(\mathbf{r}-\mathbf{r}_i) \quad \textbf{(11)}$$

where ε(r) is the grid value from the water probe at position $\mathbf{r}$, ΔV is the associated volume element of that grid point, and the a_{ij} represent the expansion coefficients with indices i and j as defined above. The arrangement of a particular surface patch and the determination of the distance $|\mathbf{r}$-$\mathbf{r}_i|$ and angle ϑ of a given point in space is illustrated in Figure 2.

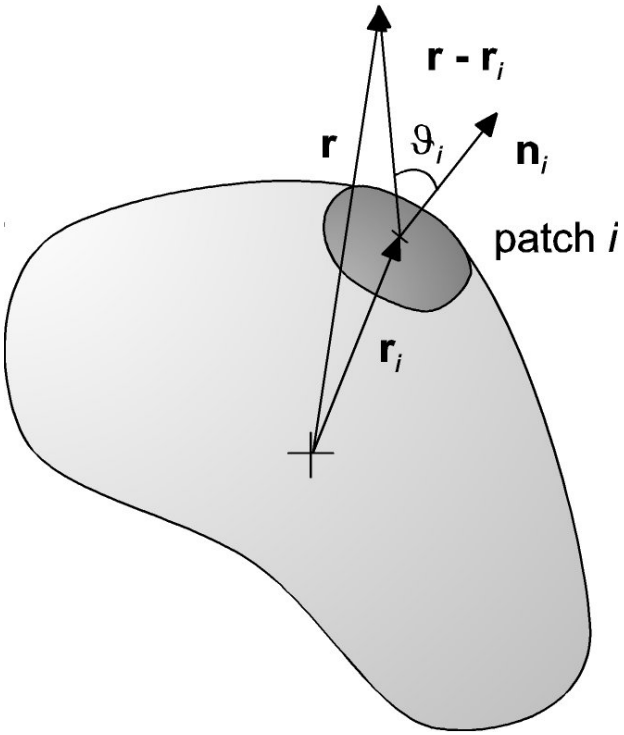

Figure 2: Schematic representation of the vectors corresponding to a surface patch and the basis functions. Also indicated is the normal vector of a patch *i*, n_i, and the angle ϑ_i between normal vector and distance vector $\mathbf{r}$-$\mathbf{r}_i$.

The interaction part of the free energy is then given by integrating eq. 11

$$\begin{aligned}\Delta G_{\text{int}} &= \int_{V_{\text{int}}} \rho_{\text{int}}(\mathbf{r})\, dr = \alpha \int_{V_{\text{int}}} \sum_i \sum_j a_{ij} f_j(\mathbf{r}-\mathbf{r}_i)\, d\mathbf{r} \\ &= \alpha \sum_i \sum_j a_{ij} \int_{V_{\text{int}}} f_j(\mathbf{r}-\mathbf{r}_i)\, d\mathbf{r}\end{aligned} \quad . \tag{12}$$

The integral is solved analytically yielding a "structural" constant c_j for each order of the Legendre polynomials *j*, thus simplifying the oben equation to

$$\Delta G_{\text{int}} = \int_{V_{\text{int}}} \rho_{\text{int}}(\mathbf{r})\, dr = \alpha \sum_j c_j \sum_i a_{ij} \; . \tag{13}$$

The hydration free energy then becomes

$$\Delta G_{\text{hyd}} = \alpha \sum_j c_j \sum_i a_{ij} + \beta S + \gamma V \; . \tag{14}$$

4.5.1 Parameterization

The implementation of the compression strategy as described above requires a sequence of three steps: (i) the three-dimensional structure of the molecule in the PDB-format,[28] (ii) the corresponding molecular surface and its division into patches, and (iii) the appropriate molecular interaction field for each molecule.

4.5.1.1 Three-dimensional structures

The three dimensional structures of the 81 molecules were created with the SYBYL program package.[29] The molecular editor was used to build the structures and an energetically favorable conformation was calculated based on SYBYL's internal force field.

4.5.1.2 Molecular surface and patch generation

In the second step, for each molecule the solvent accessible surface and volume were computed[26] and each molecular surface was divided into a certain number of surface patches. A "nearest atom" algorithm is applied to accomplish the assignment of surface patches. This algorithm measures the distance of a particular surface point to all atoms of the molecule, subtracts the van der Waals radius[30] of the atom, and assigns the PDB serial number of the atom being closest to the surface point. Such an assignment accounts for the stronger interaction of surface exposed atoms with the solvent. Thus, the set of points belonging to the same atom forms a patch on the surface. A resulting patch distribution is illustrated in Figure 3. For clarity only three hydrogen atoms and one carbon atom are considered.

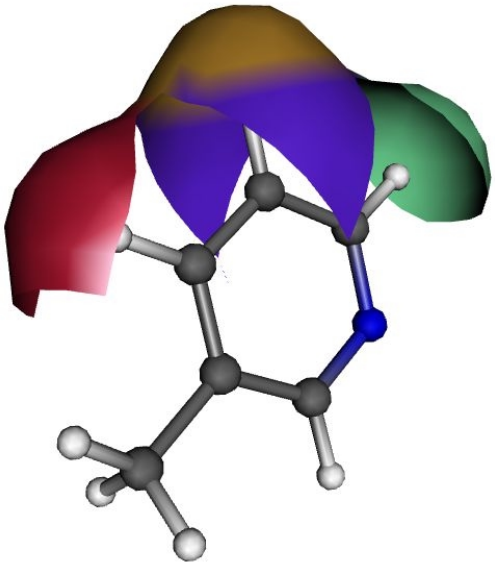

Figure 3: Patch distribution of 3-methylpyridine. The green, red, and brown patches belong to the underlying hydrogen atoms, the two blue patches correspond to the carbon atom in 5-position.

It becomes necessary to label the patches in a consistent manner to build a regression model for the interaction part of the 3D-FED as indicated in eq. 11. Therefore, the atom serial number from the PDB file was used as a distinctive feature for the patch assignment on the molecular surface. Each patch is now associated with a particular atom of the molecule, which is then combined with a classification scheme based on the atomic structure of the molecule. The atom associated with a particular patch is identified to be a certain structural atom type so that there are as many distinguishable surface patches as there are atom types in the classification scheme.

The structure analysis we used in our procedure follows that of Ghose et al.[31], which consists of roughly 120 atomic structure types characterized by their first and second neighbors. As one can see from Figure 3, some atoms have associated with them more than one surface patch, most often the benzene carbons, and obviously there are many separate surface patches that belong to the same atom type. To account for these situations, all patches related to the same atom type are treated as

identical so that the number of expansion coefficients, a_i per Legendre function j in eq. 11 is limited to the number of different atomic structure types

$$\frac{\varepsilon(\mathbf{r})}{\Delta V} = \sum_{i=1}^{120} n_i \sum_j a_{ij} f_j(\mathbf{r} - \mathbf{r}_i) . \quad \textbf{(15)}$$

Again ε divided by the volume V is the energy density at position **r**, i is now the number of atomic types and n_i is the *number of occurrences* of a surface patch associated with atom type i. Finally, we get for the hydration free energy

$$\Delta G_{\text{hyd}} = \alpha \sum_{i=1}^{120} n_i \sum_j a_{ij} c_j + \beta S + \gamma V . \quad \textbf{(16)}$$

The number of different atomic types occurring in a chosen training set of molecules is multiplied by the order of the Legendre polynomial to give the total number of independent variables, i.e. the number of columns of the design matrix, for an ordinary least squares fit. The number of *dependent* variables, i.e. the number of rows of the design matrix, is determined in the next section.

4.5.2 Least squares fit

Since the number of grid points is equal to the number of dependent variables for a least squares fit, further manipulation in order to reduce the number of rows of the design matrix is advised. Cubic grids with much smaller dimensions, or *patch grids*, were used to reduce the number of grid points. The center of the patch grid has the same coordinates as the reference point of the associated surface patch. From the grid surrounding the entire molecule we consider for regression only those grid points that lie within one of the patch grids possessing attractive interaction. In addition, the angle ϑ_i between the surface normal of the patch and the vector $\mathbf{r}$-$\mathbf{r}_i$ had to be smaller than 110°. Figure 4 shows such a scenario giving an example of the arrangement of the patch grid, the patch surface, and its normal vector.

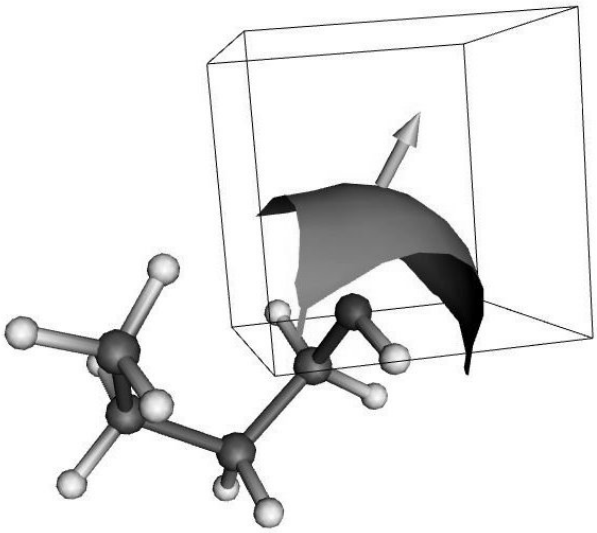

Figure 4 The volume of the patch grid, the patch surface of the oxygen of 1-butanol, and the corresponding normal vector are shown. The edges of the grid surrounding the whole molecule are omitted for clarity.

This sort of treatment reduces the average number of relevant grid points substantially, although it is clearly a function of the dimensions of the patch grids. For the set of 81 molecules the resulting optimal side length of the cubic patch grids was 3.5Å giving an average number of grid points of about 12,000 per molecule. The parameter k controlling the exponential part of the distance function (see eq. 9) was similarly optimized leading to a value of $k = 0.66\text{Å}^{-1}$. The resulting values for the structural constant c_j in eq. 13 are listed in Table 1.

The optimized values for both the size of the patch grids and the exponential decay of the distance function modeling the 3D-FED according to eq. 13 allows us to fit the a_{ij} in eq. 15 with respect to reference GRID data. Given a_{ij}, the undetermined parameters of model II (see eq. 16) are obtained by fitting α, β, γ to experimental values of ΔG_{hyd}. The results from fitting a model grid density to reference GRID data (divided by the volume element) for all 81 molecules simultaneously are summarized in Table 1. Considering the number of data points, the correlation is excellent.

Table 1: Results of the least squares fit according of the model density with respect to the reference GRID values according to model II. The correlation coefficient, the mean error, and the *F* statistic are listed depending on the maximum order of the Legendre expansion.

Order *j* of Legendre polynomial	0	1	2
c_j (Å³)	58.6593	19.2983	-6.60042
R^2	0.901	0.919	0.922
σ (kcal/mol)	0.295	0.267	0.264
F	142627	90321	62073

One can see from the slightly increasing correlation coefficient with larger *j* that the exponential distance dependence alone is able to capture a substantial part of the molecular interaction field computed by GRID. The improvement of the fit statistics by adding angle dependent terms to the mapping function is anticipated but angular terms turn out to have little influence and are almost negligible beyond *j*=1.

Consequently, the results of the ΔG_{hyd} fit are reported for *j*=1 only. The correlation coefficient is R^2 = 0.88 (*F* = 187.55) with a mean error of σ = 1.38 kcal/mol. Although both the correlation coefficient and mean error are as expected slightly better for the simple GRID model I according to eq. 8, the quality of the patch-based fit is a clear indication of the validity of the compression approach. The correlation between experimental and calculated hydration free energies is depicted in Figure 5.

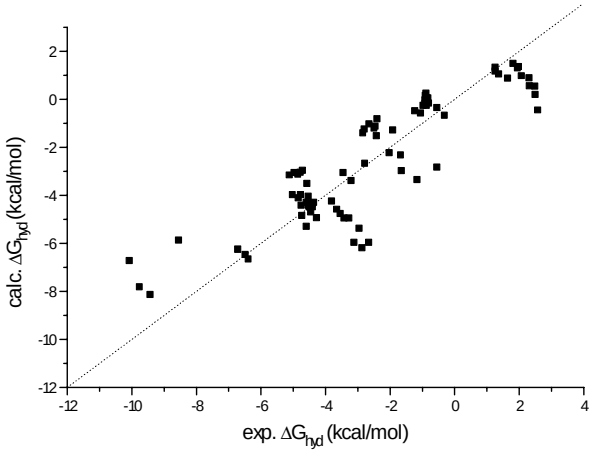

Figure 5 Calculated free energies of hydration vs. experimental as derived from the compressed model II, eq. 16, by fitting the parameters in two steps.

4.6 Final model: Simultaneous use of local and global information

The investigations described in the previous sections were used to validate our model for the hydration free energy. First, it was shown that the interaction part of the 3D-FED can be well approximated by a molecular interaction field generated from the water probe of the GRID force field. Second, modeling the approximated 3D-FED by a small amount of surface related parameters does not alter significantly the quality of the fit results.

Given the success of the simple GRID/patch model, one might ask how to further improve the correlation while retaining the predictive capabilities. On one hand, the physics of the solvation process is approximately captured in the GRID model (or its compressed representation) and the surface and volume term. On the other hand, deficiencies of the approximation are accounted for by scaling the contributions to yield experimental data in an independent step. It is therefore conceivable that the model (eq. 16) can be significantly improved if global information, i.e. experimental ΔG_{hyd}, enters the parameterization process already in the patch parameter fit. One must seek ways to use local information (GRID data) and global information (ΔG_{hyd}) *simultaneously* to determine the a_{ij} and α, β, γ of eq. 16.

We see from eq. 16 that α is nonlinearly coupled to the a_{ij}. An iteration scheme is therefore needed to find solutions of the minimization problem. We keep the functional form of eq. 16 and rearrange in such a way as to be compatible with the design matrix of model II

$$\frac{\Delta G_{\mathrm{hyd}}^{\mathrm{exp}} - \beta S - \gamma V}{\alpha} = \sum_{i=1}^{120} n_i \sum_{j=0}^{1} a_{ij}\, c_j \ . \qquad \textbf{(17)}$$

Again, the Legendre polynomial is considered up to j=1 only. The number of rows of the design matrix resulting from eq. 15 increases by the number of molecules in the training set. Thus, the total design matrix consists of two blocks, *A* and *B,* representing the grid energies divided by the volume element and the LHS of eq. 17, respectively. Although these dependent variables correspond to different physical properties, the parameter basis is the same. The design matrix is properly scaled and a weight factor, λ, is introduced to find the optimal balance between the two blocks of the matrix. This weight factor is defined in such a way that a value of λ=1 corresponds to the situation where the purely linear model (eq. 17) is used while λ=0 represents eq. 15 from model II:

$$\lambda = \frac{w}{1+w} \qquad \textbf{(18)}$$

where w is the weight factor for the block *B* of the design matrix. If both blocks are weighted equally then w=1 and λ=0.5. The function χ^2 to be minimized then is

$$\chi^2 = \sum_i \left(obs_{i,\mathrm{A,ref}} - obs_{i,\mathrm{A}}\right)^2 + w \sum_j \left(obs_{j,\mathrm{B,ref}} - obs_{j,\mathrm{B}}\right)^2 \ . \qquad \textbf{(19)}$$

The summation of the block *A* is taken over all relevant grid points *i*. The reference observable, $obs_{i,\mathrm{A,ref}}$, is the grid value of point *i* divided by the volume, whereas the modeled observable, $obs_{i,\mathrm{A}}$, is given by eq. 15. Block *B* is summed over the number of molecules in the training set *j*. Here, $obs_{i,\mathrm{B,ref}}$ and $obs_{i,\mathrm{B}}$ are the LHS and RHS of eq. 17, respectively.

4.6.1 Iteration cycle

For the initial design matrix the values of the parameters α, β, and γ from the previous least square regression according to model II serve as starting values to calculate the LHS of eq. 17. Subjecting this design matrix to a least squares algorithm, a new set of parameters a_{ij} is found. From these parameters the interaction part of the 3D-FED was calculated, inserted into eq. 16, and a new set of parameters α, β, and γ is found again by linear regression. These three parameters are used to recalculate the LHS of eq. 17 to build a new design matrix and thus closing the iterative cycle. The convergence of the mean square error of the calculated free energies of hydration of the training set was used as a stopping criterion. To estimate the predictive capability of the model we used a test set of 10 molecules with known experimental hydration free energies. Several iteration cycles were performed with varying values of λ to find the model with the best predictive power.

The value of λ corresponding to the minimum deviation is 2.44×10^{-4}. Upon further reduction, the fit results approach the same statistics (e.g. *rms* error, correlation coefficient, etc.) as obtained from model II where only the grid values were considered for the linear regression. This behavior is consistent with our expectations and assures the validity of the iterative approach.

4.6.2 Results

Only 30 of the possible 120 atom types are present in the training set for which the parameter values are reported in Table 2. Since the Legendre expansion was considered up to j=1, each atom type has two associated parameters.

Table 2: Atom types according to the classification scheme by Ghose et al.[31] and the parameters of the final model for j=0 and j=1 terms in the Legendre expansion.

Atom type (i)	a_{i0} [kcal mol^{-1} Å^{-3}]	a_{i1} [kcal mol^{-1} Å^{-3}]
1	0,598	1,555
2	-1,880	2,311
3	-1,124	3,827
5	-18,968	16,393
6	-1,395	-9,582
8	-0,156	-6,085
15	-3,978	5,015

Atom type (i)	a_{i0} [kcal mol^{-1} Å^{-3}]	a_{i1} [kcal mol^{-1} Å^{-3}]
16	-4,211	0,903
24	-4,500	1,200
25	-10,471	6,820
27	-2,220	8,069
36	-32,929	7,481
38	-25,772	45,315
39	-2,680	24,452
40	-26,808	44,656
46	-2,668	2,245
47	-6,114	4,122
48	-14,899	-0,274
49	10,920	1,274
50	-41,451	15,919
51	-7,080	5,887
52	-6,155	6,537
56	-122,558	93,539
57	-17,818	-41,028
58	-111,788	46,367
59	-43,361	-5,998
60	40,258	-58,330
66	-57,875	20,781
72	-71,096	24,861
75	-93,476	-13,826

The corresponding c_j values are reported in Table 1. The final model equation together with the values for the parameters α, β, and γ is then

$$\Delta G_{\text{hyd}} = 7.99 \cdot 10^{-4} \sum_{i,j} c_j a_{ij} + \left(\frac{0.06176}{\text{Å}^2} S - \frac{0.04083}{\text{Å}^3} V \right) \text{kcal/mol} \,. \qquad \textbf{(20)}$$

The correlation coefficient obtained for the training set is $R^2 = 0.99$ and the *rms* error turns out to be $\sigma = 0.27$ kcal/mol. Calculated hydration free energies are plotted against the experimental values in Figure 6.

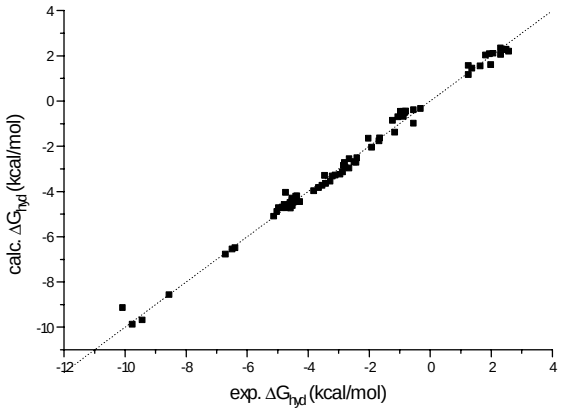

Figure 6 Values of ΔG_{hyd} obtained from the iterative procedure are plotted against experimental data.

Since we used the same reference data as No et al.,[14] it is legitimate to assess the quality of our model by a direct comparison of the *rms* error. No et al.[14] reported value of 0.43 kcal/mol is slightly higher than our value of 0.27 kcal/mol. Since both approaches use a 3D-FED as statistical basis, and both use different formulations for the empirical model, the concept of the 3D-FED and surface-compressed forms thereof is certainly valid.

For all training set members the sum $\beta S + \gamma V$ is always positive. This is another confirmation of the validity of our approach because the sum directly measures the cavity contribution. The individual surface and volume terms have no apparent physical meaning.

The mean square error for predictions of ΔG_{hyd} of the test set is $\sigma = 0.63$ kcal/mol thus being within range of experimental uncertainty ensuring the models predictive capability. The calculated free energies of hydration for the test set together with the experimental values are listed in Table 3.

Table 3: Experimental[15] and calculated hydration free energies from the final model (in kcal/mol) for the test set.

Compound	exp. ΔG_{hyd}	calc. ΔG_{hyd}
1-hexene	1.73	1.50
2-methylpentane	2.56	2.29
3-hexanol	-3.73	-4.30
4-ethylpyridine	-4.66	-4.68
cis-1,2-dimethylcyclohexane	1.60	1.68
ethylpropionate	-2.83	-3.26
hexanal	-2.85	-4.30
isopentylacetat	-2.24	-2.91
methylformate	-2.82	-2.21
tert-butylbenzene	-0.44	-0.54

Table 3 shows excellent agreement of calculated and experimental hydration free energies except for the value of the hexanal molecule. The poor prediction of ΔG_{hyd} for hexanal can be attributed to the dominant conformation or the conformation mixture of hexanal in solution which might be very different to the one generated by the SYBYL force field. Similar predictive restrictions due to an increase of conformational degrees of freedom were also obtained by No et al.[14]

Two further aspects of our model should be discussed. One is the assignment of surface patches, since it represents a degree of freedom for the entire parameterization procedure. Our initial approach uses a pure distance criterion, i.e. each surface point is associated with the closest atom. More elaborate assignment strategies could allow contributions to a surface element by several atoms, and therefore might change the results. Secondly, the parameterization quality strongly depends on the quality of the underlying GRID force fields. Since GRID has proved its value in many applications where the focus was on energetic considerations, more accurate estimates of solvent entropy terms could further improve the empirical model.

A promising approach towards a chemical understanding of solvation phenomena would be the identification of relevant surface patches. With a given patch distribution application of principal component or partial least squares algorithms[20] offers a promising route towards a further condensed representation of surface patches. Inspection of the underlying molecular structure of the newly

formed patches is expected to render identification of structural fragments relevant to the solvation process. Work in this direction is currently underway.

4.7 Conclusions

An empirical model was presented to allow for the fast prediction of free energies of hydration. The statistical basis of this model was derived from the concept of a three-dimensional free energy density which is accessible through both explicit and implicit solvent models. In our approach the interaction part of the 3D-FED is approximated by an appropriate molecular interaction field generated by the program GRID. In a second step, the molecular interaction field is modeled as a linear combination of surface functions. A standard atom classification scheme was used to associate the surface functions to particular atom types and to establish a model based on group contributions.

The final model was parameterized with respect to the simultaneous prediction of local grid data as well as global hydration free energies. The resulting rms deviation for a training set of 81 molecules is 0.27 kcal/mol and for a test set of 10 molecules 0.63 kcal/mol. Our use of local and global data represents a significant departure from other empirical work. Because our approach retains a maximum of physical information, its predictive capability is optimzed. Although the main focus of this work was the description of the model and its parameterization, it is desirable to extend the training set towards a complete coverage of all structural types of the atom classification scheme is desired.

4.8 Literature

1) Kollmann, P. A. *Acc. Chem. Res.* **1996**, *29*, 461.
2) Straatsma, T. P. *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH: New York, 1996, Vol. 9; Chapter 2.
3) King, P. M. Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications; Escom Science Publishers: Leiden, The Netherlands, 1993, pp. 267-314.
4) W. L. Jorgensen *Encyclopedia of Computational Chemistry*, P. v.R. Schleyer, Ed.; Wiley: New York, 1998; Vol. 2.
5) Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2000**, *122*, 2878.
6) Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys*. **1999**, *110*, 1329.
7) Best, S. A.; Merz Jr., K. M.; Reynolds, C. H. *J. Phys. Chem B* **1999**, *103*, 714.
8) Roux, B.; Simonson, T. *Biophysical Chemistry* **1999**, *78*, 1.
9) Papazyan, A.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 11254.
10) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
11) Tomasi, J.; Persico, M. *Chem. Rev* **1994**, *94*, 2027.
12) Luque, F. J.; Barril, X.; Orozco, M. *J. Comput.-Aided Mol. Design* **1999**, *13*, 139.
13) Oprea, T. I.; Waller, C. L. *Reviews in Computational Chemistry*, Vol. 11; Wiley-VCH: New York, 1997; Chapter 3.
14) No, K.T.; Kim, S.G.; Cho, K.-H.; Scheraga, H.A. *Biophysical Chemistry* **1999**, *78*, 127.
15) Viswanadhan, V. N.; Ghose, A. K.; Singh, U. C.; Wendoloski, J. J. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 405.
16) Böhm, H.-J.; Klebe, G.; Kubinyi, H. *Wirkstoffdesign*; Spektrum, Akad. Verl.: Heidelberg, 1996; Chapter 21.
17) Greco, G.; Novellino, E.; Martin, Y. C. *Reviews in Computational Chemistry*, Vol. 11; Wiley-VCH: New York, 1997; Chapter 4.
18) Cramer, D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc*. **1988**, *110*, 5959.
19) Goodford, P. *J. Chemometrics* **1996**, *10*, 107.
20) Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, *185*, 1.
21) Baroni, M; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. *Quant. Struct.-Act. Relat*. **1993**, *12*, 9.
22) Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. *J. Mol. Struct*. (*Theochem*) **2000**, *503*, 17.
23) Alifrangis, L. H.; Christensen, I.T.; Berglund, A.; Sandberg, M.; Hovgaard, L.; Frokjaer, S. *J. Med. Chem.* **2000**, *43*, 103.
24) Böhm, H.-J.; Klebe, G.; Kubinyi, H. *Wirkstoffdesign*; Spektrum, Akad. Verl.: Heidelberg, 1996; Chapter 25.
25) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, 1990; Chapter 6.7.
26) Connolly, M. *Science* **1983**, *221*, 709.
27) Ben-Naim, A.; Lovett, R. *J. Phys. Chem. B* **1997**, *101*, 10535.

28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.
29) SYBYL 6.6, Tripos Inc., 1699 South Hanely Road, St. Louis, Missouri, 63144, USA.
30) Bondi, A.; *J.Phys.Chem.* **1964**; *68*; 441.
31) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762.

5 Parameterization strategy for the MolFESD concept: quantitative surface representation of local hydrophobicity

5.1 Abstract

We derive a new model for the established concept of the *molecular free energy surface density* (MolFESD) [Jäger, R.; Schmidt, F.; Schilling, B.; Brickmann, J. *J. Comput.-Aided Mol. Design* **2000**, *14*, 631]. The model parameterization makes efficient use of both local and global information about solvation thermodynamics, as formulated earlier for the problem of predicting free energies of hydration [Jäger, R.; Kast, S. M. *J. Mol. Graph. Model.*, submitted]. The free energy of transfer is separated into an interaction contribution and a term related to the cavity formation. Interaction and cavity components are obtained from the statistical three-dimensional free energy density and a linear combination of surface and volume terms, respectively. An appropriate molecular interaction field generated by the program *Grid* is used as an approximate representation of the interaction part of the three-dimensional free energy density. We further compress the three-dimensional density by means of a linear combination of localized surface functions allowing for the derivation of local hydrophobic contributions in the form of a free energy surface density. For a set of 400 compounds our model yields significant correlation (R^2=0.95, σ=0.57) between experimental and calculated log *P* values. The final model is applied to establish a correlation between partial free energies of transfer for a series of sucrose derivatives and their relative sweetness, as studied earlier in the group of the authors. We find considerable improvement regarding the *rms* error of the regression thus validating the presented approach.

5.2 Introduction

Comprehension of the hydrophobic effect[1] on both microscopic and macroscopic scales has drawn tremendous attention because of its implication to a

variety of physiological processes, i.e. metabolism, bioavailability, drug action, or receptor binding. There is an ongoing discussion in the literature concerning the molecular origin of the hydrophobic effect in terms of hydrophobic hydration and hydrophobic interaction.[2-7] A general theory on the molecular level to explain the positive free energy of transferring a nonpolar solute from the gas phase into water is yet to be conceived.

A quantitative, macroscopic treatment of the hydrophobic effect is accomplished using liquid-liquid systems and measuring the distribution behavior of a particular compound between an aqueous and an organic, immiscible phase. 1-Octanol has become the standard solvent in such a partitioning experiment[8] and the logarithm of the partition coefficient, log *P*, is the most popular descriptor in many quantitative structure-activity relationships (QSAR).[9] The availability of an extensive amount of experimental log *P* values[10] has triggered the development of numerous empirical models designed to predict the log *P* value of a given compound.[11-12] These models, based on a two-dimensional representation of the molecular structure, fail when incorporated into studies of ligand-receptor interactions. As a consequence, use of molecular fields with respect to electrostatic and steric properties derived from the three-dimensional (3D) structure of the molecule, opened the field of the more successful 3D-QSAR[13] and 4D-QSAR,[14] which in addition considers conformational aspects. Besides steric and electrostatic fields considerable attention was directed towards the establishment of other fields.[15] In particular the generation of hydrophobic fields, for instance by the program HINT,[16] has drawn broad interest since hydrophobic interactions play an important role in the binding process of a ligand and a receptor. Multiplication of atomic constants, derived from 1-octanol/water partition coefficients, with a particular distance function yields such hydrophobic fields. This spatial representation of lipophilicty remains on a qualitative level and renders inadequate for the calculation of log *P* values. Carrupt and coworkers present an excellent review of current computational approaches to lipophilicity including methodological aspects, limitations, and areas of application.[17]

To capture the physics of local hydrophobic interaction, the concept of the molecular free energy surface density (MolFESD) has been developed in the group of the authors.[18,19] Based on previously determined atomic lipophilicity parameters,[20] it was shown that the transfer free energy from one solvent to another can be represented by a surface integral of a scalar quantity, the MolFESD, over the solvent

accessible surface of that molecule.[19] This concept was successfully applied to a series of sucrose derivatives by correlating the surface integral of partial surfaces to the relative sweetness of this series of compounds.[19]

The local hydrophobicity models developed so far make no explicit use of the microscopic physics of local solvation.[13,17] In the present work, a new strategy parameterization for the MolFESD is presented that accounts for local and global information simultaneously. *Local* contributions are considered by defining the statistical *three-dimensional free energy density*[21,22] (3D-FED) as the objective function for a surface based regression model. The rigorous derivation of the 3D-FED from an exact thermodynamic, statistical basis[23] allows the plausible interpretation of model parameters in terms of local contributions. *Global* information, e.g. the log *P* value or the transfer free energy for the 1-octanol/water system, is used with an optimal weight in the parameterization process, leaving the parameter basis unchanged. The optimal regression model leads to the treatment of the transfer free energy as a surface function. This allows the local representation of lipophilicity on the molecular surface in a quantitative manner so that the surface integral over this density yields the transfer free energy. Based on our results, in a final step we obtain partial free energies of transfer from incremental surface areas of a series of sucrose derivatives. These energies are correlated to the relative sweetness of this series of compounds and the results are compared to a correlation established from an earlier model.[19]

The newly derived model is a generalization of earlier work on the prediction of free energies of hydration,[21] termed "ΔG_{hyd} model" in what follows. The theoretical basis is therefore given only briefly and the reader is referred to ref. 21 for conceptual and technical details. Beforehand, we give a short review of the MolFESD concept and its main aspects. For details, the reader is referred to ref. 18 and 19.

5.3 The MolFESD Concept

To determine the free energy of solvation of a given compound in water two main contributions have to be considered: (a) the interaction energy between the solute and the surrounding solvent and (b) entropic contributions which are mainly a

function of shape and size of the cavity which accommodates the solute. Consequently, one can distinguish two classes of molecules that constitute limiting cases in terms of their solubility in water: First, the class of ionic or strong polar compounds, where the solution thermodynamics are dominated by the electrostatic energy and second, non polar compounds, where the positive free energy of solvation (unfavorable solubility) is primarily influenced from the work done on cavity formation.

In our previous work we showed, that the electrostatic interaction energy, for dilute solutions, can be represented strictly as a function of the molecular surface.[19] Moreover, following the argumentation of many other authors, we arrived at a surface-based description for the free energy of cavity formation.[19] From these results we postulated that for all intermediate cases the free energy of solvation can be adequately described by a suitably chosen, continuous function, $\rho^{\mathrm{FESD}}(\mathbf{r}_S)$, of surface coordinates $\mathbf{r}_S$ in such a way as the integral over the entire surface S recovers the solvation free energy

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{e}} + \Delta G_{\mathrm{n}} = \oint_S \rho^{FESD}(\mathbf{r}_S)\, d\mathbf{r}_S \qquad \textbf{(1)}$$

where ΔG_e and ΔG_n are the electrostatic contributions to the free energy of solvation and the free energy of cavity formation, respectively. Having established this physical basis, we derived an empirical model for the transfer free energy. Here, the MolFESD, $\rho^{\mathrm{FESD}}(\mathrm{r}_S)$, represents the transfer free energy per unit surface and has its physical meaning only on the molecular surface. Following a similar notion, Luque et al. yields the solvation free energy from a summation of discrete surface elements.[24]

In the previous paper[19] it is demonstrated that the MolFESD can be represented as a superposition of local contributions

$$\rho^{\mathrm{FESD}}(\mathbf{r}_s) = \sum_i \rho_i(\mathbf{r}_s)\,; \quad \rho_i(\mathbf{r}_s) = F_i \mu_i(\mathbf{r}_s) \qquad \textbf{(2)}$$

where ρ_i is the MolFESD of an atomic or molecular increment i, F_i is an increment constant, and μ_i is a membership function which governs the influence of an

increment *i* to the surface. The set of surface-related parameters F_i was determined based on partial $\log P$ values from an atom classification scheme.[20] With this approach, we were able to assess the hydrophobic contribution of a given surface element in a quantitative manner. This was successfully demonstrated by correlating partial transfer free energies (obtained from incremental surface areas) to the relative sweetness of a series of sucrose derivatives.[19]

In this paper we apply a parameterization strategy to arrive at an improved model for the MolFESD. This strategy has proved useful for the prediction of hydration free energies[21] and it is now extended to build a model for a surface-based description of the transfer free energy. Essential aspects of this strategy are briefly reviewed in what follows.

5.4 Model I: The *Grid* approximation to the 3D-FED

As in the ΔG_{hyd} model the solvation free energy ΔG_{solv} is separated into an interaction and a cavity term

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{int}} \tag{3}$$

where the subscripts cav/int denote cavity and interaction contributions, respectively. Applied to the transfer free energy $\Delta G_{\mathrm{trans}}$ for the 1-octanol/water system (superscript *oct* and *water*, respectively), we get

$$\Delta G_{\mathrm{trans}} = \Delta G_{\mathrm{solv}}^{\mathrm{water}} - \Delta G_{\mathrm{solv}}^{\mathrm{oct}} = \left(\Delta G_{\mathrm{int}}^{\mathrm{water}} - \Delta G_{\mathrm{int}}^{\mathrm{oct}}\right) + \left(\Delta G_{\mathrm{cav}}^{\mathrm{water}} - \Delta G_{\mathrm{cav}}^{\mathrm{oct}}\right). \tag{4}$$

The transfer free energy or its proportional quantity, the log *P* value, is related to the 3D-FED:

$$\begin{aligned} \Delta G_{\mathrm{trans}} &= -2.303RT \log P = \int_V \left(\rho_{\mathrm{solv}}^{\mathrm{water}}(\mathbf{r}) - \rho_{\mathrm{solv}}^{\mathrm{oct}}(\mathbf{r})\right) d\mathbf{r} \\ &= \int_V \rho_{\mathrm{trans}}(\mathbf{r})\, d\mathbf{r} \end{aligned} \tag{5}$$

where ρ_{trans} is the 3D-FED, **r** is a spatial vector, V is the total volume of the system, R is the gas constant, and T the temperature. In analogy to eq. 4, ρ_{trans} can be separated in similar manner by redefining integration boundaries. The solvent accessible surface[25] is a reasonable border dividing the total volume into cavity and interaction regions. In the following the term *molecular surface* is taken as the solvent accessible surface described by Conolly.[25]

In the ΔG_{hyd} model[21] the interaction part of the 3D-FED was approximated with the program *Grid*[26] using the water probe only. Modeling the interaction part of the *transfer* free energy this becomes the difference of the water and the 1-octanol interaction terms as in eq. 4. Regarding the 1-octanol phase as a simple nonpolar solvent, neglecting its partially polar properties due to the hydroxyl group and the ability to dissolve a substantial amount of water,[27,28] the interaction part is best approximated by the *hydrophobic probe* supplied by *Grid*. Although characteristic features of the 1-octanol phase are ignored, the hydrophobic probe offers the best model within the *Grid* approximation. Thus, the interaction parts of the 3D-FED of water and 1-octanol are approximated by *Grid* using the water and the hydrophobic probe, respectively, as

$$\Delta G_{\text{int}} = \int_{V_{\text{int}}} \left(\rho_{\text{int}}^{\text{water}}(\mathbf{r}) - \rho_{\text{int}}^{\text{oct}}(\mathbf{r}) \right) d\mathbf{r} \approx \alpha \sum_i \left(\varepsilon_i^{\text{water}} - \varepsilon_i^{\text{phob}} \right) \qquad \textbf{(6)}$$

where ε_i is the energy value of grid point i of the water and the hydrophobic probe (superscript *phob*) and α is a model parameter.

The cavity contributions are assumed to be a linear function of a surface and a volume term.[22,29] While the issue of surface and/or volume proportionality of the transfer free energy remains a subject of active debate,[30,31] both terms where included to avoid a biased cavity model

$$\begin{aligned} \Delta G_{\text{cav}}^{\text{water}} - \Delta G_{\text{cav}}^{\text{oct}} &= \left(\beta^{\text{water}} - \beta^{\text{oct}} \right) S + \left(\gamma^{\text{water}} - \gamma^{\text{oct}} \right) V \\ &= \beta S + \gamma V \end{aligned} \qquad \textbf{(7)}$$

with S being the molecular surface of the water phase, V is the volume enclosed by this surface, and β and γ are the parameters. In this equation we implicitly assume

that the molecular surface is similar in both phases and the difference of surface and volume parameters is described by a single parameter only. A constant term in eq. 7 is neglected for reasons discussed in ref. 21. Equations 6 and 7 are combined to establish model I for partition coefficient log *P*

$$\begin{aligned} \log P &= \alpha \sum_i \left(\varepsilon_i^{\text{water}} - \varepsilon_i^{\text{hydro}} \right) + \beta S + \gamma V \\ &= \alpha \sum_i \varepsilon_i^{\text{diff}} + \beta S + \gamma V \end{aligned} \qquad \textbf{(8)}$$

where superscript *diff* denotes the difference in grid energy densities.

The clearance of the grid was set to 4Å and the spacing of grid points was chosen to be 0.5Å. Such grid dimensions result in an average number of grid points of approximately 34,400. For each grid point, the energy value from the hydrophobic probe was subtracted from the energy value of the water probe. This difference according to eq. 6 divided by the volume element of the grid gives the approximation of the interaction part of the 3D-FED. The molecular surface and volume were computed for each molecule,[25] and for a set of 400 compounds with known experimental log *P* value the parameters α, β, and γ were determined by linear least squares regression. Despite the simplicity of the model concerning the approximation of the interaction part of the 3D-FED for the transfer free energy, the fit results, R^2=0.81 with a *rms* error of σ=1.64 kcal/mol are good.

The same regression was then repeated omitting the volume term. Interestingly, the fit statistics remain almost identical (R^2=0.80, σ=1.67 kcal/mol) to the former fit, where the volume term was part of the regression model. This leads to the assumption that volume proportional terms play a negligible role for the transfer free energy. In contrast, leaving the surface proportional term out of the regression model reduced the quality of the fit significantly (R^2=0.73, σ=1.94 kcal/mol). From these results, we conclude that the transfer free energy is dominated by contributions proportional to the molecular surface. We further discussion this issue below.

Next, following the outline of the ΔG_{hyd} model derivation,[21] the reference grid is approximated by a linear combination of surface-based functions. This is an intermediate step towards the final model where local and global thermodynamic information will simultaneously enter the regression model.

5.5 Model II: Expansion of the 3D-FED in surface-based functions

The rational behind this second model is twofold: one is the improvement of computational speed, the other is the establishment of a set of structure related parameters that allow the fast prediction of $\log P$ values by means of a group contribution approach. Furthermore, since the model functions are located on the molecular surface, the surface representation of local lipophilicity in a quantitative manner becomes possible. This aspect will be further examined after the final model section.

The essential characteristic of this second model is the approximation of the reference FED from *Grid* by a linear expansion of surface-based functions $\varphi(\mathbf{r}\text{-}\mathbf{r}_i)$. These functions were chosen as the product of an exponential distance dependence and angular dependent terms, which are given by Legendre polynomials with *j* being the degree of the expansion. The vectors $\mathbf{r}$ and $\mathbf{r}_i$ point from the origin of the coordinate system to a particular grid point and to a given reference point of patch *i*, respectively. The angle is determined from the normal vector of the reference point of patch *i* and $\mathbf{r}\text{-}\mathbf{r}_i$. Also, a parameter *k* was introduced to govern the exponential behavior and was set to an optimized value of 0.8Å^{-1} which was derived from preliminary investigations. The linear combination to approximate the *Grid* FED therefore reads

$$\rho_{\text{int}}^{\text{diff}}(\mathbf{r}) = \alpha \frac{\varepsilon^{\text{diff}}(\mathbf{r})}{\Delta V} \approx \alpha \sum_i \sum_j a_{ij} \varphi_j(\mathbf{r} - \mathbf{r}_i) \qquad \textbf{(9)}$$

where ΔV is the volume element occupied by a single grid point and α is a model parameter. For the evaluation of this equation certain preliminary steps are omitted here, that are described in detail elsewhere.[21] For clarification, a brief summary of these preceding steps is given here:

For each molecule of the training set the molecular surface was computed.[25] The surface was then divided into a certain number of patches *i*, each patch representing the surface area of the underlying atom. An atom classification scheme[20] was employed so that each patch is not only associated with the closest atom but also with the appropriate type of that atom from the classification scheme.

Since the classification scheme uses roughly 120 different atom types available, the number of different surface patch types is also restricted

$$\frac{\varepsilon(\mathbf{r})}{\Delta V} = \sum_{i=1}^{120} n_i \sum_j a_{ij} \varphi_j (\mathbf{r} - \mathbf{r}_i) \qquad \textbf{(10)}$$

where i is now the index of atom types and n_i is the *number of occurrences* of a surface patch associated with atom type i.

Integrating eq. 10 gives the interaction part of the transfer free energy. Analytical integration yields a "structural constant", c, for each expansion degree j of the Legendre polynomial. With the cavity contribution from eq. 7, model II for log P value therefore reads

$$\log P = \alpha \sum_{i=1}^{120} n_i \sum_j a_{ij} c_j + \beta S + \gamma V \; . \qquad \textbf{(11)}$$

The parameterization is accomplished in two steps. First, we determine the parameters a_{ij} by linear least squares regression according to eq. 10. The enormous amount of data points can be substantially reduced, since only a fraction of all grid points is relevant in order to capture the interaction of a chosen probe and the target structure. For that purpose, a cubic sub grid, or *patch grid*, was assigned to each surface patch enclosing only those grid points that are in close neighborhood to the surface patch and therefore being most influential. Again, details of such reduction strategies are described elsewhere.[21] For the given training set of 400 compounds the optimal side length of the patch grids was found to be 2.5Å.

The parameters α, β, and γ were subsequently fitted by linear regression. The results for the first step are reported in Table 1 for different degrees of expansion of the Legendre polynomial.

Table 1 Results of the least squares fit of the model density with respect to the reference *Grid* values according to model II (eq. 10). The correlation coefficient R^2, the mean error σ, and the *F* statistic are listed depending on the maximum order of the Legendre expansion.

Order *j* of Legendre polynomial	0	1	2
c_j [Å³]	32.9381	10.8363	-3.7062
R^2	0.882	0.890	0.893
σ [kcal/mol]	0.623	0.602	0.596
F	213201	116423	79724

The excellent approximation of the grid energy density by the linear combination of surface-based functions is shown in Figure 1. The parameters a_{ij} are taken from the j=1 expansion. The yellow solid surface represents a constant energy density calculated from *Grid*. The green, mesh-like surface corresponds to the same energy density modeled from the surface functions.

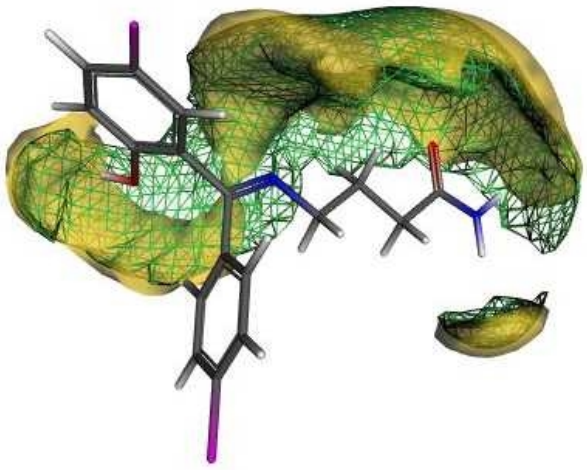

Figure 1 Overlap of two iso-surfaces for the molecule progabide. The yellow surface represents a constant density value from the *Grid* calculations. The green surface is the result for the same density modeled from the linear combination of the surface values a_{ij} (see eq. 10).

Although the size of the training set is five times greater than the one used to predict free energies of hydration,[21] the correlation coefficients are very similar. Also, since the average size of the molecules is larger in the $\log P$ data set, the amount of data points of the regression model exceeds a factor of five. Taking this into consideration the correlation is excellent. As previously observed,[21] the exponential distance function alone (j=0) captures a substantial part of the reference data. Adding angular terms yields an expected improvement, although it becomes insignificant beyond j=1 with respect to the accuracy of the model.

Consequently, the fit results for the free energy of transfer are reported for the j=1 expansion only. According to eq. 11 we obtain a correlation coefficient of R^2=0.83 with σ=1.54 kcal/mol. The slight improvement when compared to the simple *Grid* model I might not be overrated. We conclude that the energy density determined by *Grid* can be well modeled by an expansion in terms of Legendre polynomials according to eq. 10 yielding comparable fit statistics to model I.

The observations made in model I regarding the omission of the volume term are again confirmed. The fit statistics are only marginally affected when the volume term is left out of regression model II. Since both models show this invariance to a volume proportional term, for the final model described in the next section, the cavity contribution is assumed to be a function of the surface only.

5.6 Final model: Simultaneous use of local and global information

While in model II the parameterization was performed in two distinct steps, rearrangement of eq. 11 allows the global information to enter the regression model together with the local grid data to find an optimal set of parameters a_{ij}. Omitting the volume term eq. 11 is rewritten

$$\frac{\log P_{\text{exp}} - \beta S}{\alpha} = \sum_{i=1}^{120} n_i \sum_{j=0}^{1} a_{ij}\, c_j \qquad \textbf{(12)}$$

where $\log P_{\text{exp}}$ is the experimental $\log P$ value of a given substance of the training set. All other variables have been defined above. The initial design matrix therefore consists of two blocks, termed block *A* and block *B*. The parameter basis is equal for

both blocks as one can easily see by comparing eqs. 10 and 12. The dependent variables for block *A* are the energy values computed from *Grid* divided by the volume element, whereas block *B* uses the LHS of eq. 12 as reference data. With such a matrix scheme the function χ^2 to be minimized is given by eq. 13

$$\chi^2 = \sum_i \left(obs_{i,\mathrm{A,ref}} - obs_{i,\mathrm{A}}\right)^2 + w\sum_j \left(obs_{j,\mathrm{B,ref}} - obs_{j,\mathrm{B}}\right)^2 . \qquad \textbf{(13)}$$

The summation of block *A* is carried out over all relevant grid points *i*. The reference data, $\mathrm{obs}_{i,A,\mathrm{ref}}$, and the modeled data, $obs_{i,A}$, are given by the LHS and RHS of eq. 10, respectively. Based on eq. 12 the *B* block is built in analogous manner. Since the number of grid points *i* is large compared to the number of compounds *j* (*i* >> *j*), a weight factor *w* is introduced to find the optimal balance between the two blocks.

5.6.1 Iteration cycle

To initially compute the LHS of eq. 12 the parameters α and β from model II are used as starting values. The function χ^2 is minimized (see eq. 13) yielding a set of parameters a_{ij}. From these parameters new values of α and β are determined by least squares regression according to model II (see eq. 11) omitting the volume term. This allows to recalculate the LHS of eq. 12 and to repeat the parameterization procedure in an iterative manner until convergence is achieved.

5.6.2 Results

As before, the results for the iterative procedure are reported for the *j*=1 expansion only. The convergence of the *rms* error for the prediction of the experimental log *P* values was taken as a stopping criterion for the iteration cycle. We obtain a minimum *rms* error for the log *P* prediction of σ=0.57 log units (or 0.78 kcal/mol) while the corresponding correlation coefficient is R^2=0.95. The training set of 400 molecules comprised 89 atom types from the atom classification scheme.[20] For these atom types, the resulting parameters a_{ij} are listed in Table 2.

Table 2 Atom types according to the classification scheme by Ghose et al.[20] and the parameters of the final model for j=0 and j=1 terms in the Legendre expansion.

Atom type (i)	a_{i0} [kcal mol^{-1} Å^{-3}]	a_{i1} [kcal mol^{-1} Å^{-3}]
1	53,409	-49,752
2	14,208	-12,501
3	26,689	-26,843
5	19,038	-1,378
6	-9,692	17,823
7	-0,014	12,107
8	-15,172	3,902
9	-34,025	7,138
11	-337,530	266,667
12	725,400	-912,779
13	-43,737	-25,158
15	114,764	-72,910
16	22,988	-27,783
17	7,122	-14,064
18	-147,715	156,477
19	2,519	-37,144
20	70,520	-69,428
21	-184,355	164,391
22	196,768	-197,963
24	35,717	-46,227
25	15,407	-37,921
26	1,307	-28,504
27	-29,789	23,790
28	17,022	-34,449
29	24,566	-38,275
30	-172,451	124,232
31	-69,566	57,977
32	102,599	-127,133
33	-142,321	133,190
34	-38,530	2,676
35	3,326	-32,210
36	39,013	-24,129
37	-29,267	67,043
38	15,997	-28,475
39	-32,394	42,356
40	-19,191	7,663
41	81,500	-75,874
42	-253,751	196,069
43	-172,380	194,056
44	-112,645	93,604
46	-5,733	-7,634
47	-52,775	39,041
48	-46,753	37,020
49	-28,639	36,060
50	-50,621	1,494
51	-66,325	41,462
52	-14,056	-1,525
53	-17,787	3,503
54	-4,510	-14,266

Atom type (i)	a_{i0} [kcal mol^{-1} Å^{-3}]	a_{i1} [kcal mol^{-1} Å^{-3}]
56	-98,561	66,538
57	-13,460	-67,202
58	-104,299	27,255
59	-19,112	-39,154
60	37,136	-91,475
61	-89,267	17,870
62	-219,804	66,224
66	-312,465	236,463
67	-73,084	10,594
68	97,552	-208,446
69	-125,206	99,179
70	-9,784	14,268
71	34,449	13,419
72	-54,458	42,981
73	97,987	-90,401
74	-60,895	-37,753
75	-80,600	-17,668
76	98,618	-133,236
77	-5,376	20,874
78	199,837	-271,567
79	-85,309	58,111
81	27,729	-102,977
83	91,825	-108,728
84	-24,330	-26,998
86	87,878	-83,598
87	-31,514	27,565
88	1,574	6,254
89	58,509	-49,160
90	1,551	-1,719
91	74,386	-74,935
94	109,924	-94,889
95	-2,585	0,234
96	152,798	-169,690
99	28,582	-44,374
106	22,699	-9,884
107	84,576	-94,481
108	-272,909	214,790
109	-167,539	136,562
110	13,258	-86,145
117	-31,768	-46,726

We obtain for the final model the following equation

$$\log P = 1.754 \cdot 10^{-4}\ \mathrm{mol\,kcal^{-1}} \sum_{i=1}^{120} n_i \sum_{j=0}^{1} a_{ij}\, c_j + 0.019598\ \mathrm{Å^{-2}}\, S\ . \qquad \textbf{(14)}$$

By fitting local and global information simultaneously, we were able to lower the *rms* error by a factor of two compared to the simple model I thereby validating our approach. We might anticipate further improvement through a better model of the 1-octanol phase beyond the *Grid* approximation. In addition, refinement of the cavity contributions is advisable. Based on the conclusions of several authors,[6,30,32] enhancement is to be expected from incorporation of descriptors related to shape and curvature of the molecular surface. Nevertheless, the *rms* error for the transfer free energy is in agreement or slightly better when compared to other fragment approaches.[11,20] The resulting correlation plot is shown in Figure 2, where calculated log *P* values are plotted against experimental log *P* data.

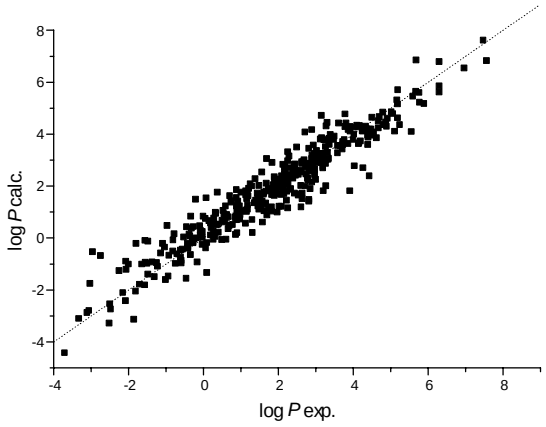

Figure 2 Log *P* values as obtained from our final model (see eq. 14) are plotted versus experimental data. The fit statistics are R^2=0.95 and σ=0.57 log units.

Table 3 compares the experimental log *P* values of the 400 compounds present in the training set to the calculated ones from eq. 14.

Table 3 Experimental and calculated log *P* values as obtained from the final model according to eq. 14.

Compound	log P_{exp}	log P_{calc}
1-bromopropane	2,10	2,18
17-alpha-hydroxyprogesterone	3,17	3,01
17-methyltestosterone	3,36	3,82
2-chloroaniline	1,90	1,87
2-chloropyridine	1,22	1,89
2-decanone	3,73	3,58
2-iodoaniline	2,32	1,85
2-naphthylamine	2,28	2,30
2-nitrophenol	1,79	1,83
3-chloroaniline	1,88	1,90
3-nitrophenol	2,00	1,90
4-chloroaniline	1,88	1,90
4-methoxyacetophenone	2,03	2,17
4-nitrophenol	1,91	1,90
6-azauridine	-2,14	-2,12
acebutolol	1,71	0,87
acetazolamide	-0,26	0,42
acetylsalicylate	1,19	1,09
aciclovir	-1,56	-1,82
alclofenac	2,47	2,58
alpha-bromotoluene	2,92	2,95
alprazolam	2,12	1,99
alprenolol	2,89	2,61
altretamine	2,73	1,98
amantadine	2,44	1,71
amben	-1,55	-0,11
aminocaproic acid	-2,95	-0,55
aminosalicylic acid	1,32	0,19
amiodarone	7,57	6,81
amitriptyline	4,92	4,59
amoxicillin	-1,99	-1,02
ampicillin	-1,13	-0,60
anthracene	4,45	4,17
aprobarbital	1,37	1,34
ascorbic acid	-1,64	-1,02
atenolol	0,16	0,65
atropine	1,83	1,66
azatadine	3,59	3,26
azelaic acid	1,57	1,37
baclofen	-0,96	0,46
bendroflumethiazid	1,19	1,68
benorylate	2,15	1,90
benznidazole	0,91	0,18
benzotrifluoride	3,01	3,03
betamethasone	2,01	2,14
betaxolol	2,81	2,85
biperiden	4,25	4,11

Compound	log P_{exp}	log P_{calc}
bisoprolol	1,87	2,03
bromazepam	1,69	2,09
bromoethane	1,61	1,67
bromomethane	1,19	1,26
bromopride	2,83	1,46
brotizolam	2,79	2,79
bunitrolol	1,91	2,23
bupranolol	2,97	3,14
buprenorphine	3,29	4,29
busulfan	-0,52	0,02
caffeine	-0,07	0,28
carazolol	3,59	3,55
carbamazepine	2,19	2,73
carbenicillin	1,13	0,55
carbon tetrachloride	2,83	2,57
carbutamide	1,01	1,01
carfecillin	2,96	2,59
carmustine	1,53	1,55
carvedilol	4,11	4,32
cathine	0,67	0,34
cefacetrile	-0,45	-1,57
cefaclor	-1,79	-0,23
cefadroxil	-2,06	-0,91
cefaloram	0,20	0,10
cefalotine	-0,41	-0,22
cefamandole	-1,47	-0,14
cefatrizine	-2,24	-1,27
cefazolin	-0,58	-0,88
cefmetazole	-0,60	-0,49
cefoperazone	-0,74	-1,00
cefotaxime	-1,36	-0,94
cefuroxime	-0,16	-0,94
cetirizine	1,70	3,04
chloralhydrate	0,99	0,88
chloramphenicol	1,14	1,06
chloroquine	4,63	3,84
chlorothiazide	-0,24	-0,05
chlorpheniramine	3,17	3,84
chlorpromazine	5,19	4,61
chlorpropamide	2,27	2,12
chlorprothixene	5,18	5,30
cimetidine	0,40	0,51
cinnarizine	5,77	5,58
ciprofloxacin	-1,08	-0,23
clindamycin	2,16	1,69
clobazam	2,12	2,07
clofazimine	7,48	7,60
clomethiazole	2,12	2,24

Compound	log P_{exp}	log P_{calc}
clomipramine	5,19	5,14
clonazepam	2,41	2,61
clonidine	1,57	2,52
clotiazepam	3,49	3,77
cloxacillin	2,43	2,81
cocaine	2,30	2,38
codeine	1,14	1,80
colchicine	1,30	1,98
corticosterone	1,94	1,88
cortisol	1,61	1,29
cortisone	1,47	1,22
coumarin	1,39	2,27
cromolyn	1,92	1,76
cyclophosphamide	0,63	0,88
cysteine	-2,49	-2,55
cytarabine	-2,51	-3,30
dapsone	0,97	1,22
debrisoquine	0,75	1,36
desipramine	4,90	4,28
dexamethasone	2,01	2,24
dexfenfluramine	3,36	3,59
dextroamphetamine	1,76	1,10
diamorphine	1,14	1,65
diazepam	2,99	2,92
diazoxide	1,20	1,60
diclofenac	4,40	3,58
dicloxacillin	2,91	3,18
didanosine	-1,24	-0,97
diethylstilbestrol	5,07	4,77
diflunisal	4,44	2,37
digitoxin	2,83	2,69
digoxin	1,26	2,07
diltiazem	2,80	2,22
diphenhydramine	3,27	3,50
disulfiram	3,88	4,16
dosulepine	4,49	4,17
doxorubicin	0,10	-1,35
enoxacin	-0,91	-0,68
enprofylline	0,33	0,97
ephedrine	0,93	1,40
erythromycin	2,54	1,64
estradiol	4,01	3,77
estriol	2,45	2,83
estrone	3,13	3,79
ethanol	-0,31	0,36
ethinylestradiol	3,67	3,53
etoposide	0,60	1,21
fampridine	0,32	0,05
fenbufen	3,20	2,68
fenfluramine	3,36	3,59
finasteride	3,03	3,35
fleroxacin	-0,55	-0,45
flucloxacillin	2,61	2,57

Compound	log P_{exp}	log P_{calc}
fludrocortisone	1,67	1,32
flufenamic acid	5,25	4,35
flunarizine	5,78	5,21
flunitrazepam	2,06	1,97
fluorouracil	-1,00	-1,62
fluoxetine	3,82	4,41
flupentixol	4,51	4,63
fluphenazine	4,36	4,21
flurbiprofen	4,16	3,64
flutamide	3,35	3,05
ftorafur	-0,27	-0,16
furosemide	2,03	1,20
ganciclovir	-2,07	-2,43
gitoformate	3,16	2,89
gitoxin	1,68	1,79
glipizide	1,91	1,71
griseofulvin	2,18	2,69
haloperidol	3,23	3,76
hexobarbital	1,49	1,46
histidine	-3,32	-3,12
hydrochlorothiazide	-0,07	0,09
hydroflumethiazide	0,36	0,50
hydroxyurea	-1,80	-2,06
hymecromone	1,90	2,34
ibuprofen	3,50	3,90
ifosfamide	0,86	0,77
imipramine	4,80	4,83
indinavir	2,92	2,89
indometacin	4,27	2,69
indoprofen	2,77	2,74
iodoethane	2,00	2,03
iodomethane	1,51	1,62
iohexol	-3,05	-2,81
iopromide	-2,05	-1,23
isocarboxazid	1,49	2,00
isoniazid	-0,70	-0,63
isosorbide	1,31	0,69
isosorbide-2-mononitrate	-0,40	-0,42
isosorbide-5-mononitrate	-0,40	-0,35
isotretinoin	6,30	5,84
isoxicam	2,83	1,40
ketamine	2,18	2,46
ketanserin	3,29	3,00
ketoconazole	4,34	4,28
lacidipine	5,56	4,08
lamivudine	-0,93	-1,48
lanatosid c	0,07	-0,40
levamisole	1,84	2,89
levobunolol	2,40	2,28
levodopa	-2,74	-0,70
levofloxacin	-0,28	-0,09
lidocaine	2,26	2,96
lidoflazine	5,60	5,44

Compound	log P_{exp}	log P_{calc}
lincomycin	0,20	0,56
lomefloxacin	-0,80	-0,09
loratadine	5,20	5,69
lorazepam	2,51	2,51
lovastatin	4,26	4,09
m-245-trichlorophenol	3,72	3,40
m-246-tribromophenol	4,13	4,25
m-iodoaniline	2,86	1,88
m-xylene	3,20	3,24
mannitol	-3,10	-2,89
mebendazole	2,83	3,43
medazepam	4,41	3,86
mefenamic acid	5,12	4,09
mefoxine	-0,02	-0,26
mefruside	1,54	1,13
meloxicam	3,01	2,24
memantine	3,28	1,99
mephenesin	1,41	1,44
mepindolol	2,30	2,69
meprobamate	0,70	0,46
mercaptopurine	-0,18	-0,01
methadone	3,93	4,09
methamphetamine	2,07	2,17
methaqualone	2,50	3,48
methimazole	-0,34	-0,66
methotrimeprazine	4,68	4,49
methylphenobarbital	1,84	1,80
metipranolol	2,28	2,42
metoclopramide	2,62	1,18
metoprolol	1,88	1,66
metronidazole	-0,02	0,01
mexiletine	2,15	1,79
midazolam	3,27	2,85
moricizine	2,98	3,23
morphine	0,76	1,42
moxalactam	-0,58	-0,98
n-heptylamine	2,57	1,67
n-octylamine	2,90	2,14
nadolol	0,71	1,61
nalidixic acid	1,59	0,59
naloxone	2,09	1,66
naproxen	3,34	3,01
nebivolol	4,04	2,76
nedocromil	2,22	2,15
niacinamide	-0,37	0,13
nicotine	1,17	1,49
niflumic acid	4,43	4,04
nimesulide	2,60	2,52
nitrazepam	2,25	2,20
nitromethane	-0,35	0,06
nordazepam	2,93	2,68
norethindrone	2,97	3,06
norfloxacin	-1,03	-0,37
nortriptyline	4,04	4,25
o-xylene	3,12	3,18
octan-1-ol	3,00	3,24
ofloxacin	-0,28	-0,09
omeprazole	2,23	1,44
ornidazole	0,60	0,73
orphenadrine	3,77	3,76
ouabain	-1,70	-1,80
oxacillin	2,38	2,34
oxamniquine	2,24	1,78
oxazepam	2,24	2,08
oxprenolol	2,10	2,37
oxyphenbutazone	2,72	4,06
p-iodoaniline	2,34	1,88
p-xylene	3,15	3,24
papaverine	2,95	3,56
paracetamol	0,51	0,66
paraldehyde	0,67	0,80
pefloxacin	0,27	0,24
penbutolol	4,15	3,62
penicillin_g	1,83	1,02
penicillin_v	2,09	1,20
penta-1-4-dien	2,47	2,44
pentazocine	3,31	4,43
pentobarbital	2,10	1,99
pentoxifylline	0,29	0,75
perphenazine	4,20	3,72
pethidine	2,45	2,63
phenacemide	0,87	0,12
phenacetin	1,58	1,40
phenazone	0,38	1,75
phenobarbital	1,47	1,55
phenprocoumon	3,62	4,41
phentermine	1,90	1,62
phenylbutazone	3,16	4,70
phenytoin	2,47	2,54
picricacid	0,89	1,84
pimozide	6,30	6,77
pindolol	1,75	1,88
pipamperone	2,02	1,68
piperacillin	0,50	-0,08
piperazine	-1,50	-1,00
piracetam	-1,54	-0,95
pirenzepine	0,10	-0,20
piretanide	3,92	1,80
piroxicam	1,98	0,97
practolol	0,79	0,50
prazepam	3,73	3,69
prednisolone	1,62	1,73
prednisone	1,47	1,53
primidone	0,91	1,06
probenecid	3,21	1,80
procainamide	0,88	0,83

Compound	log P_{exp}	log P_{calc}
procarbazine	0,06	0,73
progabide	2,97	2,97
progesterone	3,87	3,62
proguanil	2,53	2,53
promazine	4,55	4,06
promethazine	4,81	4,35
propicillin	2,65	1,71
propofol	3,79	4,27
propoxyphene	4,18	4,31
propranolol	2,98	2,64
protirelin	-2,47	-2,75
proxyphylline	-0,77	0,14
pyrazinamide	-0,60	-0,76
pyrene	4,88	4,44
pyrilamine	2,98	2,42
pyrimethamine	2,69	2,93
quazepam	4,03	3,98
quinidine	2,64	2,92
quinine	2,64	2,85
ranitidine	0,27	-0,06
remoxipride	2,13	2,82
retinol	5,68	5,63
ribavirin	-1,85	-3,15
rifampicin	1,87	1,77
ropivacaine	2,90	4,15
saccharin	0,91	0,76
salicylate	2,26	1,14
secnidazole	0,22	0,39
secobarbital	1,97	2,05
selegiline	2,90	3,04
simvastin	4,68	4,22
sotalol	-0,44	0,41
spironolactone	2,26	3,30
stavudine	-0,81	-0,52
sucrose	-3,70	-4,44
sulfadimethoxine	1,63	1,03
sulfaguanidine	-1,22	-1,11
sulfameter	0,41	0,20
sulfamethizole	0,54	1,04
sulfamethoxazole	0,89	1,21
sulfaperine	0,34	0,93
sulfapyridine	0,00	0,82
sulfinpyrazone	2,30	3,14
sulfisomidine	-0,33	0,82
sulfisoxazole	1,01	1,74
sulindac	3,05	2,88
sulphadiazine	-0,09	0,45
sulphadimidine	0,89	1,54
temafloxacin	-0,20	1,47
temazepam	2,19	2,04
terbutaline	0,08	1,53
terfenadine	5,69	6,83
testosterone	3,32	3,27
tetrachloroethene	3,40	3,96
tetracycline	-1,47	-1,41
tetrahydrocannabinol	6,97	6,53
tetrazepam	3,20	2,73
tetroxoprim	0,56	0,91
theophylline	-0,02	0,23
thiamphenicol	-0,27	-0,21
thioanisole	2,74	2,74
thioridazine	5,90	5,16
tiabendazol	2,47	2,39
tiapride	0,90	0,68
timolol	1,83	1,18
tinidazole	0,23	0,60
tolazamide	1,45	1,71
tolazolin	2,65	1,50
tolbutamide	2,34	2,36
toliprolol	1,93	2,25
tolmetin	2,79	2,30
tramadol	2,63	2,57
tranylcypromine	1,49	1,08
tretinoin	6,30	5,60
triamcinolone	1,16	1,00
triamterene	0,98	0,70
triazolam	2,42	2,52
trichlormethiazide	0,56	0,60
trichloroethene	2,61	2,00
trichloromethane	1,97	2,20
trifluoperazine	5,03	4,84
trihexyphenidyl	4,49	3,94
trimeprazine	4,71	4,65
trimethoprim	0,91	1,08
trimetrexate	2,55	2,43
triprolidine	3,92	4,27
valproate	2,75	2,44
verapamil	3,79	4,76
warfarin	2,70	2,99
xamoterol	0,61	-0,07
xylose	-3,02	-1,77
zalcitabine	-1,30	-1,51
zofenopril	4,40	3,94

5.7 Surface representation of the final model by means of a molecular free energy surface density

Despite the success of the MolFESD approach in terms of a quantitative description of local hydrophobicity,[19] a more rigorous derivation of model parameters is desired. Because of its sound physical basis, the corresponding three-dimensional quantity, i.e. the 3D-FED, offers a convenient starting point to derive a free energy surface density. For a given expansion degree *j* of the Legendre polynomials, the summation over *j* in eq. 11 can be carried out yielding a parameter f_i

$$f_i = \sum_j a_{ij} c_j \hat{=} \left\langle \rho^{FESD} \right\rangle_i S_i \qquad \textbf{(15)}$$

which has the meaning of an average FED, $<\rho^{FESD}>$, when divided by the area S_i of the associated patch *i*. For the sake of clarity, *i* is again taken as the index of surfaces patches of a molecule. Obviously, if two patches are associated with the same atom type, the f_i value is identical.

Such a surface-based representation is advantageous when partial surfaces are considered and one is interested in their contribution to global observables, e.g. the transfer free energy. The treatment of partial surface areas is already possible using our final model with the condensed description of the interaction part in terms of f_i values (see eq. 15). The summation in eq. 14 is to be taken over all patches present for the given partial surface, with *S* being the related surface area, yielding a partial thermodynamic quantity. When fractions of patches exist, the corresponding fraction of the f_i values has to be taken into account.

Fractional handling is avoided when a *continuous* surface representation is employed. Toward a description in terms of a free energy surface density, we accomplish a continuous representation of the patch-based f_i values by a weighted, normalized summation for each surface point *s*

$$f_s = \frac{\omega \sum_i f_i g_i(d_{is})}{\sum_i g_i(d_{is})} \qquad \textbf{(16)}$$

with

$$g(d_{is}) = e^{-ad_{is}^2} \tag{17}$$

where ω is a scaling factor, f_s is the interaction contribution of a surface point *s*, $g_i(d_{is})$ is a Gaussian-type function with parameter *a*, and d_{is} is the distance *along the surface* between a given surface point, *s*, and the reference point of patch *i*. While in principle the summation in eq. 16 is carried out over all patches *i*, we consider only the neighboring patch values f_i to contribute to a given surface point. The model equation for the transfer free energy can now be written, without any loss of global information, as

$$\Delta G_{\text{trans}} = \alpha \sum_s f_s + \beta S \ . \tag{18}$$

5.7.1 Surface representation

While the f_i values from eq. 15 were localized on the reference points of the surfaces patches, we have now available a continuous description of the transfer free energy in terms of contributions from all points of the molecular surface. Under the condition that the sum of all surface point areas O_s recovers the total surface area *S*, eq. 18 is written as

$$\Delta G_{\text{trans}} = \alpha \sum_s \frac{f_s}{O_s} O_s + \beta \sum_s O_s \ . \tag{19}$$

From eq. 19 follows immediately

$$\Delta G_{\text{trans}} = \sum_s (\alpha \ f_s' + \beta) O_s \ ; \quad f_s' = \frac{f_s}{O_s} . \tag{20}$$

The relation to the free energy surface density becomes now apparent in eq. 21

$$\Delta G_{\text{trans}} = \oint_S \rho^{FESD}(\mathbf{r})\,d\mathbf{r} \approx \sum_s \rho^{FESD}(\mathbf{r}_s)O_s = \sum_s (\alpha\, f_s' + \beta)O_s \qquad \textbf{(21)}$$

where s is the index of all surface points, $\mathbf{r}_s$ are the corresponding point coordinates, and O_s is the point surface area.

While there is in principle no need for a surface-based representation of the parameters derived from the final model, such a representation is advantageous when partial surfaces are considered. If the free energy surface density is known for a molecule, integration of a surface fraction yields a partial contribution to the global observable, e.g. the transfer free energy. Since both local and global information were considered for the final model in the previous section, the patch-based f_i values represent a highly plausible distribution of local information gathered form the 3D-FED. The f_i values act as sampling points thereby preserving the local information for a continuous representation on the molecular surface (see eq. 16). Additionally, global information is completely retained as the transfer free energies from eq. 21 will be identical to the values obtained from eq. 11 omitting the volume term.

The free energy surface density as determined from eq. 21 is now applied to a series of sucrose derivatives and the results are compared to a model based on a set of incremental log P values.[19]

5.7.2 Application to the relative sweetness of a series of sucrose derivatives

A series of sucrose halogens derivatives has been examined previously with the MolFESD model.[19] It was found that the partial transfer free energy, computed from the surface area belonging to the fructose portion of the sucrose, nicely correlates with the relative sweetness of these derivatives. Those findings are based on particular assumptions regarding the model of the sweetness receptor[33-35] and the proportionality of the sweetness to the equilibrium constant K for the association of a sweetener and its receptor.[36] In this earlier treatment, the MolFESD values were parameterized based on a set of established incremental log P values[20] and a continuous surface representation was found employing a membership function satisfying eq. 1. These investigations yielded a correlation coefficient of R^2=0.94 and a *rms* error of 0.81. We will now reexamine this correlation behavior using the parameters derived from our model according to eq. 14. Both patch-based, f_i values,

and surface point-based, f_s values, are employed to establish a correlation as indicated above.

5.7.3 Patch-based approach

To obtain partial thermodynamic information from incremental surface areas, our final model equation (see eq. 14) is well suited for such a purpose. The sum in eq. 14 is to be taken over all patches that are present on the given partial surface area. In this equation, the sum of the f_i values represents the interaction part of the 3D-FED even if the sum is not taken over all patches i of a complete surface. If integration boundaries cut patches into fractions, the corresponding fraction of the f_i value has to be considered in the summation. With this approach the following results for the sucrose derivatives were obtained (Table 4). The sequence of compounds (**1**-**16**) is identical to the one used in the previous publication.[19]

Table 4 Transfer free energies obtained form the final model (see eq. 12). Summation was carried out over all patches *i* that are present on the surface area of the fructose portion. Compounds **2-6** are derivatives of sucrose (**1**), while compounds **8-16** are derivatives of 4-chloro-4-deoxy-*galacto*-sucrose (**7**).

	Substitution Positions	Surface Increment, [Å²]	$\Delta G_{transfer}$ [kcal/mol]	Relative Sweetness[19]
1	-	157,36	2,69	1
2	1'-Cl	163,56	0,88	20
3	6'-Cl	167,49	0,84	20
4	1'-Cl, 4'-Cl	171,47	-0,91	30
5	1'-Cl, 6'-Cl	173,58	-0,86	80
6	1'-Cl, 4'-Cl, 6'-Cl	181,54	-2,66	100
7	-	157,56	-1,63	5
8	6'-Cl	167,06	-0,77	50
9	1'-Cl	163,88	-0,75	120
10	4'-Cl, 6'-Cl	174,66	-2,56	160
11	1'-Cl, 4'-Cl	171,60	-2,54	220
12	1'-Cl, 6'-Cl	173,79	-2,49	600

	Substitution Positions	Surface Increment, [Å²]	$\Delta G_{transfer}$ [kcal/mol]	Relative Sweetness[19]
13	1'-6'-Cl, 4'-F	170,18	-3,35	1000
14	1'-Cl, 4'-Cl, 6'-Cl	181,52	-4,28	2200
15	1'-Cl, 6'-Cl, 4'-Br	184,90	-4,29	3000
16	1'-6'-Cl, 4'-I	188,79	-4,73	3500

Correlating the partial transfer free energy with the relative sweetness yields R^2=0.96 and σ=0.64. We consider this correlation excellent, because of its high correlation coefficient and the fact that none of the above listed compounds were present in the training set of the log *P* parameterization (except for pure sucrose). This selection of sweetners thus represents a small test set and the obtained fit results render strong confirmation of our model approach.

Since the fit quality exceeds the one from ref. 19 this is a clear indication of the plausibility of the f_i values in terms of local contributions. Such local information was already present in the parameterization process therefore leading to a more rigorous rationalization of local moieties. Assigning local or fragmental contributions from global information alone remains somewhat arbitrary, although its successful application[19] is undisputed. In Figure 3 we plot the natural logarithm of the relative sweetness versus the transfer free energy obtained from the partial surfaces of the sucrose derivatives.

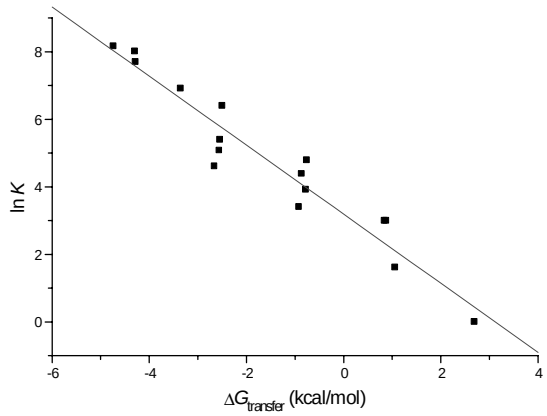

Figure 3 The relative sweetness, ln *K*, plotted as a function of the predicted partial transfer free energy, $\Delta G_{transfer}$; R^2=0.96, σ=0.64

5.7.4 Continuous representation

As it was shown above, depending on the available surface we were able to evaluate partial thermodynamic quantities from the discrete, patch-based f_i values alone. One might ask for a continuous representation in particular to benefit from visual inspection. Such a continuous function, i.e. the MolFESD (see eq. 1), can be derived from the model parameters of eq. 14. Although the local information inherent in our final model was used for the derivation of the MolFESD, an unambiguous surface representation is unattainable. According to eq. 17, the Gaussian coefficient *a* can be adjusted to a particular application.

We now focus again on the series of sucrose derivatives. We determine the best possible surface representation with respect to an optimal correlation of partial surface integrals and the relative sweetness. The Gaussian parameter *a* of eq. 17 was altered in such a way, as the turning point of the Gaussian function would increase from 1Å to 4Å in integer steps. $\rho^{FESD}(\mathbf{r}_s)$ was determined according to eq. 21 and integration of the partial surface area of the fructose portion was performed yielding the related partial transfer free energy. Results are summarized in Table 5.

Table 5 Resulting partial transfer free energies (reported in kcal/mol) obtained from integration of the surface area of the fructose portion for different values of *a* of the Gaussian function in eq. 17. The ordering of compound is identical to Table 4. Correlation coefficients R^2 and *rms* errors σ for the correlation of ΔG_{trans} with the relative sweetness of compounds **1-16** (see Table 4) are also listed.

Compound	***a*=0.5 Å^{-2}**	***a*=0.125 Å^{-2}**	***a*=0.055 Å^{-2}**	***a*=0.0313 Å^{-2}**
1	13,29	13,12	13,12	13,12
2	7,53	8,27	8,73	8,84
3	6,16	6,85	7,53	7,82
4	0,17	1,88	2,91	3,25
5	0,34	1,83	2,85	3,25
6	-7,36	-4,74	-3,08	-2,51
7	-6,80	-6,80	-6,80	-6,80
8	0,39	0,16	0,22	0,28
9	2,04	1,53	1,07	0,96
10	-7,71	-6,57	-5,66	-5,32
11	-5,54	-5,03	-4,80	-4,75
12	-6,97	-6,12	-5,37	-5,09
13	-10,57	-9,60	-8,74	-8,34
14	-14,62	-12,56	-11,31	-10,91
15	-14,22	-12,45	-11,25	-10,85
16	-19,12	-15,30	-13,36	-12,79
R^2	0.956	0.962	0.960	0.958
σ *[log units]*	0.70	0.65	0.67	0.68

We chose the parameter *a* such as to go from a very localized representation (turning point of the Gaussian function at 1Å, *a*=0.5Å^{-2}) to a more delocalized picture allowing a greater influence from more distant patches. As one can see from Table 5 the correlation coefficient has its maximum when the turning point of the Gaussian function is chosen around 2Å. The resulting representation of the MolFESD on the molecular surface is presented in Figure 4. Global information is completely conserved, since surface integration of the MolFESD yields the log *P* value or the transfer free energy.

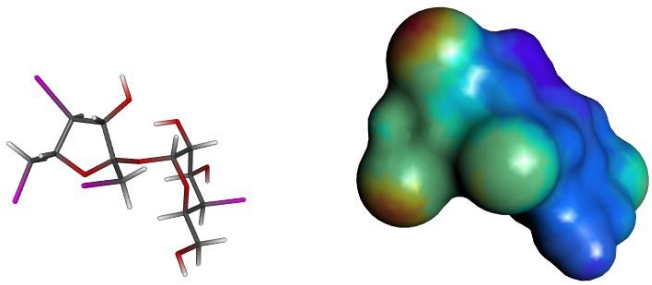

Figure 4 The molecular surface (right) of the 4,1',6'-trichloro-4,1',4',6'-tetradeoxy-4'-iodo-*galacto*-sucrose (left) is shown. Color coding is used to represent the MolFESD on the surface (yellow – positive values, blue – negative values). The Gaussian parameters for this representation is a=0.125Å^{-2}.

Since the fit quality for the continuous representation is comparable to the simple, patch-based approach, we conclude that both the continuous and discrete representations bear a consistent description of the concept of local hydrophobicity. The results obtained from both descriptions with respect to the application to sweetness recognition are in good agreement with out previous model,[19] where local information was absent for the model parameterization.

Obviously, the MolFESD gains tremendous plausibility in terms of interpreting local, fragmental contributions when constructed from our final model given in eq. 14. We are able to assess the mutual influences of neighboring substructures on the molecular surface using a normalized Gaussian function to distribute local, discrete information, i.e. the patch-based f_i values. Apparently, with the turning point of the gauss function set to 2Å, the contributions of f_i values further away than 5Å is below 5% thus yielding an optimal correlation for the application described above.

For the applied parameterization strategy the empirical character is notably reduced since the 3D-FED is principle accessible through general, statistical thermodynamic theory. In this work, we use empirical molecular interaction fields (derived from *Grid*) as a model for the 3D-FED and employ experimental data for a

proper scaling of model parameters. Finally, we arrive at a model that allows the fast and reliable prediction of log *P* values with the benefit of providing local information simultaneously. The predictive capability of this model is strong for both global and partial thermodynamic properties thereby confirming the strategy for our model derivation.

Additionally, dependence on partial, atomic log *P* values[11] has become obsolete for our approach. The independence of previously determined fragmental log *P* contributions marks a significant departure from many other empirical work. We are optimistic that our approach and its enhanced plausibility in terms of local contributions can be efficiently used in QSAR studies, in particular for rational drug design where local aspects of the binding interface are of great importance.

5.8 Conclusion

Based on the *three-dimensional free energy density* (3D-FED) a parameterization strategy for the transfer free energy was presented. This strategy has previously been established[21] providing excellent results when applied to the prediction of hydration free energies. The interaction part of the 3D-FED is approximated by an appropriate molecular interaction field generated by the program *Grid*, while cavity contributions are modeled as a linear combination of surface and volume proportional terms. The parameters for the interaction part of the 3D-FED are related to the molecular structure allowing the fast prediction of transfer free energies by means of a group contribution method. The final model equation is derived considering both local (from the 3D-FED data) and global (experimental log *P* values) thermodynamic information *simultaneously*. For a set of 400 compounds with known experimental log *P* values, a correlation was found with R^2=0.95 and a *rms* error of 0.82 kcal/mol.

The log *P* model provides a convenient starting point for a physically motivated derivation of the established MolFESD. From the 3D-FED we find surface-based parameters capturing this local information and acting as sampling points when a continuous distribution (the MolFESD) is determined on the surface. Consequently, the MolFESD concept provides the representation of local hydrophobicity on the solvent accessible surface in a quantitative manner. The

incorporation of local information to the parameterization procedure marks the essential feature of our strategy allowing for reasonable estimates of partial thermodynamic properties.

To further validate the presented model, a correlation was established between the partial free energy surface density, taken from incremental surfaces of a series of sucrose derivatives, and their relative sweetness. Comparison of the results to an earlier model showed slight improvement thereby rendering justification of the presented approach.

5.9 Literature

1) Blokzijl, W.; Engberts, J.B.F.N. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1545.
2) Scheraga, H. A. Journal of Biomolecular Structure & Dynamics **1998**, 16, 447.
3) Lazaridis, T. *J. Phys. Chem. B* **2000**, *104*, 4964.
4) Ruelle, P.; Kesselring, U. *J. Pharm. Sci.* **1998**, *87*, 987.
5) Besseling, N. A. M.; Lyklema, J. *J. Phys. Chem. B* **1997**, *101*, 7604.
6) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.
7) Hummer, G. *J. Am. Chem. Soc.* **1999**, *121*, 6299.
8) Sangster, J. Octanol-Water Partition Coeffients: Fundamentals and Physical Chemistry; John Wiley & Sons: Chichester 1997; Vol. 2.
9) Leo, A. J.; Hansch, C. Perspective in Drug Discovery and Design **1999**, 17, 1.
10) *Star list Database*; Biobyte Corporation; Pomona, CA, 1997.
11) a) Buchwald, P.; Bodor, N. *Curr. Med. Chem.* **1998**, *5*, 353 b) Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868.
12) a) Mannhold, R.; Rekker, R. F.; Sonntag, C.; Ter Laak, A. M.; Dross, K; Polymeropoulos, E. E. *J. Pharm. Sci.* **1995**, *84*, 1410 b) Mannhold, R.; Rekker, R. F. *Perspectives in Drug Discovery and Design* **2000**, *18*, 1.
13) Oprea, T. I.; Waller, C. L. *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH: New York, 1997, Vol. 11; Chapter 3.
14) Vedani, A.; McMasters, D. R.; Dobler, M. *Quant. Struct.-Act. Relat.* **2000**, *19*, 149.
15) Waller, C. L.; Kellogg, G. E. *Network Sci*. [Electronic Publications] **1996**, *2*, (URL: http://www.netsci.org/Science/Compchem/feature10.html).
16) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comput.-Aided Mol. Design* **1991**, *5*, 545.
17) Carrupt, P.-A.; Testa, B.; Gaillard, P. *Reviews in Computational Chemistry*, Wiley-VCH: New York, 1997, Vol. 11; Chapter 5.
18) Pixner, P.; Heiden, W.; Merx, H.; Möller, A.; Moeckel, G.; Brickmann, J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309.
19) Jäger, R., Schmidt, F.; Schilling, B.; Brickmann, J. *J. Comput.-Aided Mol. Design* **2000**, *14*, 631.
20) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762.
21) Jäger, R.; Kast, S. M. *J. Phys. Chem. A.*, submitted.
22) No, K.T.; Kim, S.G.; Cho, K.-H.; Scheraga, H.A. *Biophysical Chemistry* **1999**, *78*, 127.
23) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, 1990; Chapter 6.7.
24) Luque, F.J., Barril, X. and Orozco, M., *J. Comput.-Aided Mol. Design* **1999**, *13*, 139.
25) Connolly, M. *Science* **1983**, *221*, 709.
26) Goodford, P. *J. Chemometrics* **1996**, *10*, 107.

27) Best, S. A.; Merz Jr., K. M.; Reynolds, C. H. *J. Phys. Chem B* **1999**, *103*, 714.
28) DeBolt, S. E; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5316.
29) Ben-Naim, A.; Lovett, R. *J. Phys. Chem. B* **1997**, *101*, 10535.
30) Shimizu, S.; Ikeguchi, M.; Nakamura, S.; Shimizu, K. *J. Chem. Phys.* **1999**, *110*, 2971.
31) Vitha, M. F.; Carr, P. W. *J. Phys. Chem. B* **2000**, *104*, 5343.
32) Southhall. N. T.; Dill, K. A. *J. Phys. Chem. B* **2000**, *104*, 1326.
33) Shallenberger, R. S.; Acree, T. E. *Nature* **1967**, *216*, 480.
34) Kier, L. B. *J. Pharm. Sci.* **1972**, *61*, 1394.
35) Immel, S.; Kreis, U.; Lichtenthaler, F. W. *Starch* **1991**, *43*, 121.
36) Lee, C. K. Adv. Carboh. Chem. Biochem. **1987**, 45, 199.

Hilfsmittel

- Die molekularen Wechselwirkungsfelder wurden mit dem Programm *Grid* erstellt (Version 18, Molecular Discovery Ltd, West Way House, Elms Parade, Oxford OX2 9LL England). Die Berechnung erfolgte auf einer SiliconGraphics® Workstation, Indigo2 XZ, MIPS R5000, 120 Mhz.

- Zur Berechnung der Modellparameter wurde ein Intel Pentium III, 550 MHz, 1 GB RAM, eingesetzt.

- Zur Generierung der dreidimensionalen Strukturen dienten die Programme *Sybyl* (Version 6.6, Tripos Inc., 1699 South Hanely Road, St. Louis, Missouri, 63144, USA) und *Corina* (URL: http.//www2.ccc.uni-erlangen.de/software/corina/corina.html)

- Der Text dieser Arbeit wurde mit dem Textverarbeitungsprogramm *Word* erstellt (Version Word 97, Microsoft Corporation).

- Die Grafiken wurde mit Hilfe des Programms *Molcad II* generiert (Version 1.2, Dipl.-Ing. Matthias Keil, keil@pc.chemie.tu-darmstadt.de)

- Die Graphen wurden mit dem Programm *Origin* erstellt (Version 3.5, Microcal Software, Inc. One Roundhouse Plaza, Northampton, MA 01060, USA)

Robert Jäger
Seerobenstr. 24
65195 Wiesbaden

Darmstadt, den 22.11.00

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, daß ich meine Dissertation, mit Ausnahme von Kapitel 3.4 und 3.5, selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

_____

(Robert Jäger)

Lebenslauf

Robert Martin Jäger

Familienstand:	verheiratet, 1 Kind
Geburtsort:	Böblingen
Geburtsdatum:	20. März 1970
Schule:	Ernst-Leinius-Schule, Grundschule 34128 Kassel *August 1976 – Juni 1980* Heinrich-Schütz-Schule, Gymnasium 34119 Kassel *Juli 1980 – Juni 1986* Jakob-Grimm-Schule, Oberstufengymansium 34117 Kassel *Juli 1986 – Mai 1989* *Abschluß: Abitur*
Wehrdienst:	3. Fernmeldebataillon/2 34233 Fuldatal *Juni 1989-August 1990*
Studium:	Technische Universität Darmstadt 64287 Darmstadt *Oktober 1990 – Juli 1996* Fachrichtung Chemie, *Abschluß: Dipl.-Ing.* Diplomarbeit: „Empirisches Berechnungsverfahren von oberflächenbezogenen 1-Octanol/Wasser Verteilungs-koeffizienten“
Promotion:	Technische Universität Darmstadt 64287 Darmstadt *Oktober 1996 - November 2000* Institut für Physikalische Chemie, *Abschluß: Dr.-Ing.* Dissertation: „Quantification und Lokalization of Molecular Hydrophobicity“
Auslandsaufenthalt:	University of Maryland College Park, MD 20742, USA *August 1993 - Mai 1994* Graduate School, Department of Chemistry and Biochemistry